(12) INTERNATIONAL APPLICATION PUBLISHED U ATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
8 February 2007 (08.02.2007)

PCT

(10) International Publication Number
WO 2007/015178 A2

(51) International Patent Classification: Not classified

(21) International Application Number:
PCT/IB2006/003102

(22) International Filing Date: 13 July 2006 (13.07.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
10-2005-0064402 15 July 2005 (15.07.2005) KR

(71) Applicant *(for all designated States except US)*: **KOREA RESEARCH INSTITUTE OF BIOSCIENCE AND BIOTECHNOLOGY** [KR/KR]; 52, Eoeun-dong, Yuseong-gu, Daejeon 305-806 (KR).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **SOHN, Jung-Hoon** [KR/KR]; 103-506, Nuri Apt., Wolpyeong-dong, Seo-gu, Daejeon 302-791 (KR). **CHOI, Eui-Sung** [KR/KR]; 307-1703, Song Rim, Maeul Apt., Haki-dong, Daejeon 305-328 (KR). **BAE, Jung-Hoon** [KR/KR]; 539-3, Noeun-dong, Yuseong-gu, Daejeon 305-325 (KR). **SHIN, Mi-Kyung** [KR/KR]; 504-7, Noeun-dong, Yuseong-gu, Daejeon 305-768 (KR). **YOON, Sung-Sook** [KR/KR]; 319-401, Jonggunkyo Apt., Sinbong-dong, Yuseong-gu, Daejeon 305-154 (KR). **CHUN, Chang-Soo** [KR/KR]; 7-63, Mia 4-dong, Gangbuk-gu, Seoul 142-104 (KR).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Declarations under Rule 4.17:**

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

**Published:**

— *without international search report and to be republished upon receipt of that report*

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

(54) Title: LIBRARY OF TRANSLATIONAL FUSION PARTNERS FOR PRODUCING RECOMBINANT PROTEINS AND TRANSLATIONAL FUSION PARTNERS SCREENED THEREFROM

(57) Abstract: The invention relates to techniques for the rapid screening of suitable translational fusion partners (TFPs) capable of inducing secretory production of recombinant proteins, especially proteins that are difficult to produce using conventional recombinant production methods.

WO 2007/015178 A2

# LIBRARY OF TRANSLATIONAL FUSION PARTNERS FOR PRODUCING RECOMBINANT PROTEINS AND TRANSLATIONAL FUSION PARTNERS SCREENED THEREFROM

## BACKGROUND OF THE INVENTION

### Field of the Invention

[0001] This invention is in the field of recombinant protein expression. In particular, the invention relates to techniques for the rapid screening of suitable translational fusion partners (TFPs) capable of inducing secretory production of recombinant proteins, especially proteins that are difficult to produce using conventional recombinant production methods.

### Related Art

[0002] The recombinant expression of proteins of interest is a widely used procedure to produce large quantities of proteins for research purposes or for therapeutic and other commercial uses. A variety of recombinant expression systems are known in the art, including bacterial, yeast, and mammalian host cell systems, and many different proteins have been successfully produced in these systems. However, there are also many proteins that are not easily produced using available expression systems, resulting in little or no protein expression and secretion. Methods for improving the secretion of recombinantly expressed proteins, such as overexpressing secretory factors in the host cells, using fusion proteins comprising the protein of interest fused to a well-secreted protein, and adding synthetic linker sequences, have had some success with particular proteins of interest. However, no general technique has been identified that is effective for the secretory production of all proteins.

[0003] In an effort to identify secreted proteins and novel signal sequences, several signal sequence trap systems have been developed. U.S. Patent No. 6,228,590 describes a technique for screening for mammalian signal sequences by transforming reporter protein-deficient yeast with nucleic acids comprising mammalian coding sequences fused to a reporter protein and detecting cells that secrete the reporter protein. A similar system using

invertase-deficient yeast and an invertase reporter protein is disclosed in EP0907727. Yeast-based signal sequence traps have been used to identify secreted proteins from human DNA (Klein *et al., Proc. Natl. Acad. Sci. USA 93*:7108 (1996); Jacobs *et al., Gene 198*:289 (1997)), mouse DNA (Gallicioti *et al., J. Membrane Biol. 183*:175 (2001)), zebrafish DNA (Crosier *et al., Dev. Dynamics 222*:637 (2001)), *Arabidopsis* DNA (Goo *et al., Plant Mol. Biol. 41*:415 (1999)), potato DNA (Surpili *et al., Anais de Academia Brasileira de Ciencias 74*:599 (2002)), and *Candida albicans* DNA (Monteoliva *et al., Eukaryotic Cell 1*:514 (2002)). Similar trap systems have been developed using mammalian host cells (Gallicioti *et al., J. Membrane Biol. 183*:175 (2001)) and bacterial host cells (Ferguson *et al., Cancer Res. 65*:8209 (2000). Reporter proteins that have been used in signal sequence traps include invertase (Klein *et al., Proc. Natl. Acad. Sci. USA 93*:7108 (1996)), alpha-amylase (U.S. Patent No. 6,228,590), acid phosphatase (PHO5) (Surpili *et al., Anais de Academia Brasileira de Ciencias 74*:599 (2002)), and β-lactamase Ferguson *et al., Cancer Res. 65*:8209 (2000).

[0004] A method for identifying translational fusion partners (TFPs) useful for secretion of a target protein is disclosed in WO 2005/068658. The method comprises (i) obtaining a plurality of host cells transformed with a variety of vectors comprising a library of nucleic acid fragments and a target protein-encoding nucleotide sequence fused with a reporter protein-encoding nucleotide sequence, wherein the host cells are deficient in the reporter protein, and (ii) identifying a TFP library from the host cells, wherein the TFP library comprises nucleic acid fragments which individually induce the secretion of the target protein.

## SUMMARY OF THE INVENTION

[0005] The present invention relates to a rapid and efficient automatic screening method for the identification of TFPs that are effective for inducing secretion of a target protein. The invention allows any target protein to be

secreted from a host cell, including target proteins that are not expressed or expressed only at low levels using traditional recombinant expression systems.

[0006] In one embodiment, the invention relates to a method of identifying a target protein specific TFP, said method comprising:

(i) co-transforming a plurality of reporter protein deficient host cells with a plurality of linear vectors and a nucleotide sequence encoding a target protein to produce a plurality of transformed host cells,

wherein each of said linear vectors comprises a nucleic acid fragment from a library of nucleic acid fragments and a nucleotide sequence encoding a N-terminal amino acid-deleted reporter protein, and

wherein said nucleotide sequence encoding a target protein comprises, at the 3' end, a nucleotide sequence encoding the N-terminal amino acids deleted from said reporter protein in said linear vector, and at the 5' end, a linker DNA;

(ii) incubating said plurality of transformed host cells under conditions effective to allow *in vivo* recombination of said linear vectors and said nucleotide sequence encoding a target protein;

(iii) identifying a cell showing an activity of the reporter protein from the plurality of transformed host cells of (ii); and

(iv) identifying a TFP from the cell identified in (iii);

wherein said TFP comprises a nucleic acid fragment which induces the secretion of said target protein.

[0007] Another embodiment of the invention relates to a method of identifying a target protein specific TFP library, said method comprising:

(i) co-transforming a plurality of reporter protein deficient host cells with a plurality of linear vectors and a nucleotide sequence encoding a target protein to produce a plurality of transformed host cells,

wherein each of said linear vectors comprises a nucleic acid fragment from a library of nucleic acid fragments and a nucleotide sequence encoding a N-terminal amino acid-deleted reporter protein, and

wherein said nucleotide sequence encoding a target protein comprises, at the 3' end, a nucleotide sequence encoding the N-terminal amino acids deleted from said reporter protein in said linear vector, and at the 5' end, a linker DNA;

(ii) incubating said plurality of transformed host cells under conditions effective to allow *in vivo* recombination of said linear vectors and said nucleotide sequence encoding a target protein;

(iii) identifying cells showing an activity of the reporter protein from the plurality of transformed host cells of (ii); and

(iv) identifying a TFP library from the cells identified in (iii);

wherein said TFP library comprises nucleic acid fragments which individually induce the secretion of said target protein.

[0008] The invention further relates to a TFP or a library of TFPs identified by the methods of the invention.

[0009] The invention further comprises a nucleic acid fragment encoding a TFP or a library of nucleic acid fragments encoding TFPs.

[0010] The invention also includes a nucleic acid comprising a nucleotide sequence encoding a TFP and a nucleotide sequence encoding a target protein.

[0011] The invention further relates to a method of producing a target protein using a TFP of the invention.

[0012] The invention additionally relates to a linear vector comprising a nucleic acid fragment from a library of nucleic acid fragments and a nucleotide sequence encoding a N-terminal amino acid-deleted reporter protein.

[0013] The invention also comprises a plurality of reporter protein-deficient host cells transformed with the library of linear vectors and a nucleotide sequence encoding a target protein of the invention.

## BRIEF DESCRIPTION OF THE DRAWINGS/FIGURES

[0014] The above and other objects, features and advantages of the present invention will be more clearly understood from the following detailed description taken in conjunction with the accompanying drawings.

[0015] FIG. 1 shows a process for deleting the invertase gene and a pop-out process of a selectable marker.

[0016] FIG. 2 shows zymogram analysis for invertase activity (lanes 1, 2 and 3: wild-type *Saccharomyces cerevisiae* Y2805; and lanes 4, 5 and 6: invertase-deficient strain (*S. cerevisiae* Y2805Δ*suc2*).

[0017] FIG. 3 photographically shows the growth of yeast cells according to carbon sources (*SUC2*: wild-type *S. cerevisiae* Y2805; and Δ*suc2*: invertase-deficient strain (*S. cerevisiae* Y2805Δ*suc2*).

[0018] FIG. 4 shows the results of Southern blotting for the deletion of the invertase gene (lanes 1 and 2: *S. cerevisiae* Y2805 (*ura3 SUC2)*; lanes 3 and 4: *S. cerevisiae* Y2805Δ*suc2U* (*URA3*Δ*suc2*); and lanes 5 and 6: *S. cerevisiae* Y2805Δ*suc2* (*ura3*Δ*suc2*).

[0019] FIG. 5 photographically shows the growth of yeast cells containing plasmids pYGAP-SNS-SUC2, pYGAP-HSA-SUC2, and pYGAP-hIL2-SUC2, on glucose and sucrose media, respectively.

[0020] FIG. 6 shows a map of plasmid YGaINV containing multiple cloning sites for the insertion of a cDNA library between the *GAL10* promoter and the mature invertase gene.

[0021] FIG. 7 shows a map of plasmids YGaF0INV, YGaF1INV and YGaF2INV containing multiple cloning sites for the insertion of a genomic DNA library between the *GAL10* promoter and the mature invertase gene with three different reading frames.

[0022] FIG. 8 shows the process of synthesis of a cDNA library with random primer and construction of a cDNA library in the TFP selection vector YGaINV.

[0023] FIG. 9 shows the process of construction of a genomic DNA library in the TFP selection vectors YGaF0INV, YGaF1INV, and YGaF2INV.

[0024] FIG. 10 shows the plasmid map of YGadV45 containing a defective *SUC2* and subcloning of a TFP library into YGadV45.

[0025] FIG. 11 shows the procedure of TFP selection for a target gene using an invertase as a reporter from a TFP library through *in vivo* recombination.

[0026] FIG. 12 shows the procedure of TFP selection for a target gene using a double reporter, lipase and invertase, as a reporter from a TFP library through *in vivo* recombination.

[0027] FIG. 13 shows tributyrin plates containing halo forming transformants (A) A halo forming plate (YPSGA with tributyrin) directly from transformation, (B) Selected transformants showing different halo sizes in tributyrin plate.

[0028] FIG. 14 shows the procedure for the construction of 9 human IL2 expression vectors with 9 selected TFPs.

[0029] FIG. 15 shows the maps of human IL2 expression vectors (A) pYGT9-IL2, (B) pYGT13-IL2, and (C) pYGT17-IL2.

[0030] FIG. 16 shows the maps of human IL2 expression vectors (A) pYGT18-IL2, (B) pYGT19-IL2, and (C) pYGT20-IL2.

[0031] FIG. 17 shows the maps of human IL2 expression vectors (A) pYGT21-IL2, (B) pYGT25-IL2, and (C) pYGT27-IL2.

[0032] FIG. 18 shows the results of SDS-PAGE of culture supernatants of yeast cells secreting human IL2 (lane M: protein size marker; lane 1: culture supernatant of yeast cells containing pYIL-KRT1-4 (WO 2005/068658) as a control for IL2 secretion; lane 2: culture supernatant of yeast cells containing pYGT9-IL2; lane 3: culture supernatant of yeast cells containing pYGT21-IL2; lane 4: culture supernatant of yeast cells containing pYGT13-IL2; lane 5: culture supernatant of yeast cells containing pYGT17-IL2; lane 6: culture supernatant of yeast cells containing pYGT25-IL2; lane 7: culture supernatant of yeast cells containing pYGT19-IL2; lane 8: culture supernatant of yeast

cells containing pYGT18-IL2; lane 9: culture supernatant of yeast cells containing pYGT27-IL2).

[0033] FIG. 19 shows the result of SDS-PAGE of culture supernatants of 38 yeast transformants obtained from the TFP selection process for human IL32α (lane M: protein size marker; lane N: untransformed cell as a negative control; lane 1 to 38: yeast transformants).

[0034] FIG. 20 shows the results of SDS-PAGE and Western blotting of culture supernatants of yeast cells secreting human IL32α (lane M: protein size marker; lane 1: culture supernatant of yeast cells containing pYGT3-IL32α; lane 2: culture supernatant of yeast cells containing pYGT21-IL32α; lane 3: culture supernatant of yeast cells containing pYGT13-IL32α; lane 4: culture supernatant of yeast cells containing pYGT25-IL32α; lane 5: culture supernatant of yeast cells containing pYGT22-IL32α and lane 6: culture supernatant of yeast cells containing pYGT11-IL32α).

[0035] FIG. 21 shows (A) a profile for fed-batch fermentation of a recombinant yeast strain containing pYGT3-hIL32α and (B) the results of SDS-PAGE for analyzing proteins secreted into the medium according to fermentation time.

[0036] FIG. 22 shows the results of SDS-PAGE of culture supernatants of yeast cells secreting human growth hormone (lane M: protein size marker; lane N: culture supernatant of untransformed yeast cells as a negative control; lane 1: culture supernatant of yeast cells containing pYGT1-hGH, lane 2: pYGT2-hGH; lane 3: pYGT3-hGH; lane 4: pYGT4-hGH; lane 5: pYGT5-hGH; lane 6: pYGT6-hGH; lane 7: pYGT7-hGH; lane 8: pYGT8-hGH; lane 9: pYGT9-hGH; lane 10: pYGT21-hGH; lane 11: pYGT13-hGH; lane 12: pYGT25-hGH; lane 13: pYGT17-hGH; lane 14: pYGT22-hGH; lane 15: pYGT32-hGH; lane 16: pYGT19-hGH; lane 17: pYGT27-hGH; lane 18: pYGT11-hGH; lane 19: pYGT40-hGH; lane 20: pYGT43-hGH; lane 21: pYGT44-hGH.

[0037] FIG. 23 shows (A) a profile for fed-batch fermentation of a recombinant yeast strain containing pYGT18-hGH and (B) the results of SDS-PAGE for analyzing proteins secreted into the medium according to fermentation time.

[0038] FIG. 24 shows a procedure for the construction of a TFP library from selected ORFs using a unidirectional deletion method.

[0039] FIG. 25 shows the results of SDS-PAGE of culture supernatants of randomly selected yeast transformants transformed with the unidirectional-deleted TFP library constructed from the ORFs selected by BLAST search.

[0040] FIG. 26 shows the results of SDS-PAGE of culture supernatants of randomly selected yeast transformants transformed with the unidirectional-deletion TFP library constructed from 35 selected ORFs.

[0041] FIG. 27 shows the results of SDS-PAGE and Western blotting (anti-hIGF) of culture supernatants of yeast cells secreting human insulin-like growth factor (Lane M; protein size marker; lane 1: culture supernatant of yeast cells containing pYGa-MFa-hIGF; lane 2: pYGa-T1α-IGF; lane 3: pYGa-T2α-IGF; lane 4: pYGa-T3α-IGF; lane 5: pYGa-T4α-IGF).

[0042] FIG. 28 shows the results of SDS-PAGE of culture supernatants of yeast cells transformed with TFP vectors for the secretion of human caspase-1 subunit P10 (lane M: protein size marker; lane 1: culture supernatant of yeast cells with pYGT1-hP10; lane 2: pYGT2-hP10; lane 3: pYGT3-hP10; lane 4: pYGT4-hP10; lane 5: pYGT5-hP10; lane 6: pYGT6-hP10; lane 7: pYGT7-hP10; lane 8: pYGT8-hP10; lane 9: pYGT9-hP10; lane 10: pYGT21-hP10; lane 11: pYGT13-hP10; lane 12: pYGT25-hP10; lane 13: pYGT17-hP10; lane 16: pYGT22-hP10; lane 18: pYGT18-hP10; lane 19: pYGT33-hP10; lane 20: pYGT19-hP10; lane 21: pYGT27-hP10; lane 22: pYGT11-hP10; lane 24: pYGT39-hP10; lane 25: pYGT40-hP10; lane 28: pYGT43-hP10; lane 29: pYGT44-hP10; lane 32: negative control).

[0043] FIG. 29 shows the results of SDS-PAGE and Western blotting (anti-IL32) of culture supernatants of yeast cells secreting human interleukin 32 gamma (lane M: protein size marker; lane C: culture supernatant of

untransformed yeast cells as a negative control; lane 1: pYGT1-IL32γ; lane 2: pYGT2-IL32γ; lane 3: pYGT3-IL32γ; lane 4: pYGT4-IL32γ; lane 5: pYGT5-IL32γ; lane 6: pYGT6-IL32γ; lane 7: pYGT7- IL32γ; lane 8: pYGT8-IL32γ; lane 9: pYGT9-IL32γ; lane 10: pYGT21-IL32γ; lane 11: pYGT13-IL32γ; lane 12: pYGT25-IL32γ; lane 13: pYGT17-IL32γ; lane 16: pYGT22-IL32γ; lane 18: pYGT18-IL32γ; lane 19: pYGT33-IL32γ; lane 20: pYGT19-IL32γ; lane 21: pYGT27-IL32γ; lane 22: pYGT11-IL32γ; lane 24: pYGT39-IL32γ; lane 25: pYGT40-IL32γ; lane 28: pYGT43-IL32γ; lane 29: pYGT44-IL32γ; lane 33: pYGT48-IL32γ; lane 35: pYGT50-IL32γ; lane 36: pYGT51-IL32γ; lane 37: pYGT52-IL32γ; lane 39: pYGT54-IL32γ).

[0044] FIG. 30 shows the results of SDS-PAGE of culture supernatants of yeast cells secreting human interleukin-2(lane M: protein size marker; lane 1: culture supernatant of yeast cells containing YGaSW-pSUN-IL2; lane 2: YGaSW-pSED-IL2; lane 3: YGaSW-pUNK-IL2; lane 4: YGaSW-pMUC-IL2).

## DETAILED DESCRIPTION OF THE INVENTION

[0045] The present invention addresses the need for a rapid and efficient screening technique for identification of a TFP specifically applicable to a target protein for maximal secretion of the target protein. While the invention is useful to optimize the recombinant expression of any protein, it is particularly useful to enable the production of proteins that cannot be produced on a large scale and/or at low cost due to their low level of expression in known expression systems.

[0046] In one embodiment, the invention relates to a method of identifying a target protein specific TFP, said method comprising:

(i) co-transforming a plurality of reporter protein deficient host cells with a plurality of linear vectors and a nucleotide sequence encoding a target protein to produce a plurality of transformed host cells,

wherein each of said linear vectors comprises a nucleic acid fragment from a library of nucleic acid fragments and a nucleotide sequence encoding a N-terminal amino acid-deleted reporter protein, and

wherein said nucleotide sequence encoding a target protein comprises, at the 3' end, a nucleotide sequence encoding the N-terminal amino acids deleted from said reporter protein in said linear vector, and at the 5' end, a linker DNA;

(ii) incubating said plurality of transformed host cells under conditions effective to allow *in vivo* recombination of said linear vectors and said nucleotide sequence encoding a target protein;

(iii) identifying a cell showing an activity of the reporter protein from the plurality of transformed host cells of (ii); and

(iv) identifying a TFP from the cell identified in (iii);

wherein said TFP comprises a nucleic acid fragment which induces the secretion of said target protein.

[0047] Another embodiment of the invention relates to a method of identifying a target protein specific TFP library, said method comprising:

(i) co-transforming a plurality of reporter protein deficient host cells with a plurality of linear vectors and a nucleotide sequence encoding a target protein to produce a plurality of transformed host cells,

wherein each of said linear vectors comprises a nucleic acid fragment from a library of nucleic acid fragments and a nucleotide sequence encoding a N-terminal amino acid-deleted reporter protein, and

wherein said nucleotide sequence encoding a target protein comprises, at the 3' end, a nucleotide sequence encoding the N-terminal amino acids deleted from said reporter protein in said linear vector, and at the 5' end, a linker DNA;

(ii) incubating said plurality of transformed host cells under conditions effective to allow *in vivo* recombination of said linear vectors and said nucleotide sequence encoding a target protein;

(iii) identifying cells showing an activity of the reporter protein from the plurality of transformed host cells of (ii); and

(iv) identifying a TFP library from the cells identified in (iii);

wherein said TFP library comprises nucleic acid fragments which individually induce the secretion of said target protein.

[0048] The library of nucleic acid fragments may be obtained from DNA of any type, including genomic DNA, cDNA, synthetic DNA, and recombinant DNA. Nucleic acids other than DNA may also be used, including RNA and non-naturally occurring nucleic acids.

[0049] TFPs may be identified from the DNA of any eukaryotic or prokaryotic organism, including bacteria, fungi (*e.g.*, yeast), plants, and animals (e.g., mammals). Suitable bacteria include, but are not limited to *Escherichia* and *Bacillus* species. Suitable yeast include, but are not limited to *Candida*, *Debaryomyces*, *Hansenula*, *Kluyveromyces*, *Pichia*, *Schizosaccharomyces*, *Yarrowia*, *Saccharomyces*, *Schwanniomyces*, and *Arxula* species. Examples of specific species include *Candida utilis*, *Candida boidinii*, *Candida albicans*, *Kluyveromyces lactis*, *Pichia pastoris*, *Pichia stipitis*, *Schizosaccharomyces pombe*, *Saccharomyces cerevisiae*, *Hansenula polymorpha*, *Yarrowia lipolytica*, *Schwanniomyces occidentalis*, and *Arxula adeninivorans*. Other fungi that may serve as a source of DNA include, but are not limited to *Aspergillus*, *Penicillium*, *Rhizopus*, and *Trichoderma* species. Plants that may serve as a source of DNA include, but are not limited to *Arabidopsis*, maize, tobacco, and potato. Suitable animals include, but are not limited to humans, mice, rats, rabbits, dogs, cats, and monkeys.

[0050] The nucleic acid fragments may be derived from the entire genome of an organism, *e.g.*, an entire genomic or cDNA library. The fragments may also be derived from any subset of the entire genome, *e.g.*, a subtracted library or a sized library.

[0051] In one embodiment, the nucleic acid fragments are derived from a library of pre-selected candidate TFPs, *e.g.*, a library comprising TFPs that have been identified in previous screens. In a particular embodiment, the

library of pre-selected candidate TFPs is a library of core TFPs that have been identified as effective TFPs for one or more target proteins.

[0052] In another embodiment, the library of pre-selected candidate TFPs is obtained by transforming a plurality of reporter protein-deficient host cells with a variety of vectors comprising a library of nucleic acid fragments and a reporter protein-encoding nucleic acid sequence, collecting cells that grow, isolating vectors from the cells, and isolating nucleic acid fragments from the vectors, thereby obtaining a TFP library comprising the nucleic acid fragments which individually induce secretion of the reporter protein.

[0053] In a further embodiment, the library of pre-selected candidate TFPs is derived from sequences identified in a genome database by searching for (i) genes containing a pre-secretion signal homologous with those of one or more previously identified TFPs; (ii) genes comprising a secretion signal sequence; or (iii) genes encoding proteins passing through endoplasmic reticulum (*e.g.*, cell wall proteins, excretory proteins, plasma membrane proteins, vacuolar proteins, bud proteins).

[0054] In another embodiment, the library of pre-selected candidate TFPs is obtained by diversifying previously identified TFPs, *e.g.*, by unidirectional deletion, mutation, addition of functional sequences (*e.g.*, glycosylation sites) or swapping of pre- and pro-signal sequences between TFPs.

[0055] In one embodiment, the nucleic acid fragments have a size of less than 1000 base pairs, *e.g.*, less than 700, 500, or 300 base pairs. In a further embodiment, the library of nucleic acid fragments is constructed by enzymatic cleavage of the DNA, by cDNA synthesis, or by recombinant DNA technology (*e.g.*, unidirectional deletion, mutagenesis).

[0056] The linear vectors of the present invention may be any vector that is functional in the selected host cell. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA

segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. The vectors of the present invention are capable of directing the expression of genes encoding target proteins to which they are operatively linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), that serve equivalent functions.

[0057] Expression of proteins in prokaryotes may be carried out with vectors containing constitutive or inducible promoters directing the expression of the target protein-reporter protein fusion. Examples of suitable *E. coli* expression vectors include pTrc (Amrann *et al.*, *Gene 69*:301-315 (1988)) and pET (Studier *et al.*, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 60-89).

[0058] For expression in yeast cells, suitable yeast expression vectors include, but are not limited to pYepSecl (Baldari *et al.*, *EMBO J. 6*:229-234 (1987)), pMFa (Kurjan *et al.*, *Cell 30*:933-943 (1982)), pJRY88 (Schultz *et al.*, *Gene 54*:113-123 (1987)), pYES2 (Invitrogen Corporation, San Diego, Calif.), and picZ (Invitrogen Corp, San Diego, Cal.).

[0059] For expression in insect cells, baculovirus expression vectors may be used. Baculovirus vectors available for expression of proteins in cultured insect cells (*e.g.*, SF9 cells) include the pAc series (Smith *et al.*, *Mol. Cell. Biol. 3*:2156-2165 (1983)) and the pVL series (Lucklow *et al.*, *Virology 170*:31-39 (1989)).

[0060] In another embodiment, the host cells are mammalian cells and the vector is a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, *Nature 329*:840 (1987)) and pMT2PC (Kaufman *et al., EMBO J. 6*: 187-195 (1987)). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, adenovirus 2, cytomegalovirus and Simian Virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells, see, *e.g.*, Chapters 16 and 17 of Sambrook *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

[0061] Preferred vectors include, but are not limited to, plasmids, phages, cosmids, episomes, viral particles or viruses, and integratable DNA fragments (*i.e.*, fragments integratable into the host genome by homologous recombination). Preferred viral particles include, but are not limited to, adenoviruses, baculoviruses, parvoviruses, herpesviruses, poxviruses, adenoassociated viruses, Semliki Forest viruses, vaccinia viruses, and retroviruses. Preferred expression vectors include, but are not limited to, pcDNA3 (Invitrogen) and pSVL (Pharmacia Biotech). Other expression vectors include, but are not limited to, pSPORTTM vectors, pGEMTM vectors (Promega), pPROEXvectorsTM (LTI, Bethesda, MD), BluescriptTM vectors (Stratagene), pQETM vectors (Qiagen), pSE420TM (Invitrogen), and pYES2TM (Invitrogen).

[0062] In one embodiment, expression vectors are replicable DNA constructs in which a DNA sequence encoding the target protein is operably linked or connected to suitable control sequences capable of effecting the expression of the target protein in a suitable host. DNA regions are operably linked or connected when they are functionally related to each other. For example, a promoter is operably linked or connected to a coding sequence if it controls the transcription of the sequence. Amplification vectors do not require

expression control domains, but rather need only the ability to replicate in a host, usually conferred by an origin of replication, and a selection gene to facilitate recognition of transformants. The need for control sequences in the expression vector will vary depending upon the host selected and the transformation method chosen. Generally, control sequences include, but are not limited to a transcriptional promoter, enhancers, an optional operator sequence to control transcription, polyadenylation signals, a sequence encoding suitable mRNA ribosomal binding and sequences which control the termination of transcription and translation. Such regulatory sequences are described, for example, in Goeddel; GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Regulatory sequences include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, *etc.*

[0063] The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein. Preferred vectors preferably contain a promoter that is recognized by the host organism. The promoter sequences of the present invention may be prokaryotic, eukaryotic or viral. Examples of suitable prokaryotic sequences include the PR and PL promoters of bacteriophage lambda (The bacteriophage Lambda, Hershey, A. D., Ed., Cold Spring Harbor Press, Cold Spring Harbor, NY (1973), which is incorporated herein by reference in its entirety; Lambda II, Hendrix, R. W., Ed., Cold Spring Harbor Press, Cold Spring Harbor, NY (1980), which is incorporated herein by reference in its entirety); the trp, recA, heat shock, and lacZ promoters of *E. coli* and the SV40 early promoter (Benoist *et al., Nature; 290*:304-310 (1981)), which is incorporated herein by reference in its entirety).

For yeast, examples of suitable promoters include, but are not limited to GAPDH, PGK, ADH, PHO5, GAL1, and GAL10. Additional promoters include, but are not limited to, mouse mammary tumor virus, long terminal repeat of human immunodeficiency virus, maloney virus, cytomegalovirus immediate early promoter, Epstein Barr virus, Rous sarcoma virus, human actin, human myosin, human hemoglobin, human muscle creatine, and human metallothionein.

[0064] Additional regulatory sequences can also be included in preferred vectors. Examples of suitable regulatory sequences are represented by the Shine-Dalgarno sequence of the replicase gene of the phage MS-2 and of the gene cII of bacteriophage lambda.

[0065] Moreover, suitable expression vectors can include an appropriate marker that allows the screening of the transformed host cells. The transformation of the selected host is carried out using any one of the various techniques well known to the expert in the art and described in Sambrook *et al.*, supra.

[0066] An origin of replication can also be provided either by construction of the vector to include an exogenous origin or may be provided by the host cell chromosomal replication mechanism. If the vector is integrated into the host cell chromosome, the latter may be sufficient. Alternatively, rather than using vectors which contain viral origins of replication, one skilled in the art can transform mammalian cells by the method of co-transformation with a selectable marker and target protein DNA. An example of a suitable marker is dihydrofolate reductase (DHFR) or thymidine kinase (see, U.S. Patent No. 4,399,216).

[0067] Nucleotide sequences encoding the target protein may be recombined with vector DNA in accordance with conventional techniques, including blunt-ended or staggered-ended termini for ligation, restriction enzyme digestion to provide appropriate termini, filling in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and ligation with appropriate ligases. Techniques for such manipulation are disclosed by

Sambrook *et al.*, supra and are well known in the art. Methods for construction of mammalian expression vectors are disclosed in, for example, Okayama *et al.*, *Mol. Cell. Biol. 3*:280 (1983), Cosman *et al.*, *Mol. Immunol. 23*:935 (1986), Cosman *et al.*, *Nature 312*:768 (1984), EP-A-0367566, and WO 91/18982, each of which is incorporated herein by reference in its entirety.

**[0068]** The host cells used in the present invention may be any host cells know to those of skill in the art. Suitable host cells include bacterial, fungal, (*e.g.*, yeast), plant, or animal (*e.g.*, mammalian or insect) cells. Suitable yeast cells include *Candida*, *Debaryomyces*, *Hansenula*, *Kluyveromyces*, *Pichia*, *Schizosaccharomyces*, *Yarrowia*, *Saccharomyces*, *Schwanniomyces*, and *Arxula* species. Specific examples include *Candida utilis*, *Candida boidinii*, *Candida albicans*, *Kluyveromyces lactis*, *Pichia pastoris*, *Pichia stipitis*, *Schizosaccharomyces pombe*, *Saccharomyces cerevisiae*, *Hansenula polymorpha*, *Yarrowia lipolytica*, *Schwanniomyces occidentalis*, and *Arxula adeninivorans*. Other suitable fungi include *Aspergillus*, *Penicillium*, *Rhizopus*, and *Trichoderma* species. Bacteria that may be used as host cells include *Escherichia*, *Pseudomonas*, and *Bacillus* species. Suitable plant host cells include *Arabidopsis*, maize, tobacco, and potato. Animal cells include cells from humans, mice, rats, rabbits, dogs, cats, monkeys, and insects. Examples include CHO, COS 1, COS 7, BSC 1, BSC 40, BMT 10, and Sf9 cells.

**[0069]** In a particular embodiment, the host cells are yeast cells, and the nucleic acid fragments are isolated from the genome or cDNA of a yeast.

**[0070]** Polynucleotides of the invention may be introduced into the host cell as part of a circular plasmid, or as linear DNA comprising an isolated protein coding region or a viral vector. Methods for introducing DNA into the host cell that are well known and routinely practiced in the art include transformation, transfection, electroporation, nuclear injection, or fusion with carriers such as liposomes, micelles, ghost cells, and protoplasts.

[0071] Any reporter protein that is rapidly and efficiently detectable may be used in the present invention. In one embodiment, the reporter protein has an activity that can be positively selected for in order to automate the screening process. In an additional embodiment, the reporter protein is a protein that is secreted into the extracellular space, *e.g.*, invertase, sucrase, cellulase, xylanase, maltase, amylase, glucoamylase, galactosidase (*e.g.*, alpha-galactosidase beta-galactosidase, melibiase), phosphatase (*e.g.*, PHO5), beta-lactamase, lipase or protease. In a particular embodiment, the secreted protein permits a cell to grow on a particular substrate. As an example of reporter system in mammalian cell, CD2/neomycin-phosphotransferase (Ceo) gene can be used as a secretion reporter in the media containing antibiotics G418 to trap the secretion pathway genes in mouse embryonic stem cells (De-Zolt et al., Nucleic Acid Res. 34:e25 (2006)).

[0072] In one embodiment, the host cells are yeast, the reporter protein is invertase and the transformed yeast cells are selected for their ability to grow on sucrose or raffinose. In another embodiment, the host cells are yeast, the reporter protein is melibiase and the transformed yeast cells are selected for their ability to grow on melibiose. In a further embodiment, the host cells are yeast, the reporter protein is amylase (*e.g.*, an endoamylase, exoamylase, β-amylase, or glucoamylase), the yeast cells are non-amylolytic, and the transformed cells are screened for their ability to degrade starch. In an additional embodiment, the step of identifying cells showing an activity of the reporter protein occurs by using a reporter protein which provides resistance to a growth inhibitor, *e.g.*, an antibiotic. In another embodiment, the reporter protein is a protein that can be detected visually, *e.g.*, green fluorescent protein or luciferase. In one embodiment, the step of identifying cells showing an activity of the reporter protein occurs by using two or more reporter proteins, *e.g.*, lipase and invertase.

[0073] The host cells of the present invention do not exhibit reporter protein activity. In one embodiment, the host cells naturally do not express the reporter protein. In other embodiments, the gene(s) encoding the reporter

protein have been deleted in whole or in part or have been mutated such that the reporter protein is not expressed or is expressed in an inactive form. Methods for rendering a cell deficient in a particular protein are well known in the art and any such method may be used to prepare the host cells of the present invention (Sambrook *et al.*, supra). For yeast, a reporter gene deficiency can be introduced using well known gene replacement techniques (Rothstein, *Meth. Enzymol. 194*:281 (1991)).

[0074] The linear vector of the invention comprises a nucleic acid fragment and a nucleotide sequence encoding a N-terminal amino acid-deleted reporter protein. The N-terminal amino acid deletion may encompass any number of amino acids as long as the deletion is sufficient to substantially eliminate reporter protein activity. For example, the deletion may encompass about 5, 10, 15, 20, 25, 30, 35, 40, 45, or 50 or more amino acids from the N-terminus of the reporter protein.

[0075] The methods of the present invention may be used with any target protein for which there is a desire for high level recombinant expression. The target protein may be one that is being studied for research purposes or one that is being produced for commercial purposes, *e.g.*, therapeutic or industrial use. The target protein may be from any plant, animal, or microorganism, and may be naturally occurring or modified in any way, as long as it can be encoded by a nucleic acid. In one embodiment the target protein is a human protein. In another embodiment, the target protein is a cytokine, serum protein, colony stimulating factor, growth factor, hormone, or enzyme. For example, the target protein may be selected from an interleukin, coagulation factor, interferon-α, -β or -γ, granulocyte-colony stimulating factor, granulocyte macrophage-colony stimulating factor, tissue growth factor, epithelial growth factor, TGFα, TGFβ, epidermal growth factor, platelet-derived growth factor, fibroblast growth factor, follicle stimulating hormone, thyroid stimulating hormone, antidiuretic hormone, pigmentary hormone, parathyroid hormone, luteinizing hormone-releasing hormone, carbohydrate-specific enzymes, proteolytic enzymes, lipases, oxidoreductases, transferases,

hydrolases, lyases, isomerases, ligases, immunoglobulins, cytokine receptors, lactoferrin, phospholipase A2-activating protein, insulin, tumor necrosis factor, calcitonin, calcitonin gene related peptide, enkephalin, somatomedin, erythropoietin, hypothalamic releasing factor, prolactin, chorionic gonadotropin, tissue plasminogen activator, growth hormone releasing peptide, thymic humoral factor, anticancer peptides, or antibiotic peptides. Specific examples include, but are not limited to human interleukin-2, human interleukin-1β, human interleukin-6, human interleukin-32α, -32β or -32γ, Factor VII, Factor VIII, Factor IX, human serum albumin, human interferon-α, -β or -γ, human granulocyte-colony stimulating factor, human granulocyte macrophage-colony stimulating factor, human growth hormone, human platelet-derived growth factor, human basic fibroblast growth factor, human epidermal growth factor, human insulin-like growth factor, human nerve growth factor, human transforming growth factor β–1, human follicle stimulating hormone, glucose oxidase, glucodase, galactosidase, glucocerebrosidase, glucuronidase, asparaginase, arginase, arginine deaminase, peroxide dismutase, endotoxinase, catalase, chymotrypsin, uricase, adenosine diphosphatase, tyrosinase, bilirubin oxidase, bovine galactose-1-phosphate uridyltransferase, jellyfish green fluorescent protein, *Candida antarctica* lipase B, *Candida rugosa* lipase, fungal chloroperoxidase, β-galactosidase, resolvase, α-galactosidase, β-glucosidase, trehalose synthase, cyclodextrin glycosyl transferase, xylanase, phytase, human lactoferrin, human erythropoietin, human paraoxonase, human growth differentiation factor 15, human galectin-3 binding protein, human serine protease inhibitor, Kunitz type 2, human Janus kinase 2, human fms-like tyrosine kinase 3 ligand, human YM1 & 2, human CEMI, human diacylglycerol acyltransferase, human leptin, human mL259, human proteinase 3, human lysozyme, human DEAD box protein 41, human etoposide induced protein 24, mouse caspase1, bovine angiogenin, and earthworm lumbrokinase.

[0076] In one embodiment, the target protein is a protein that is difficult to produce using conventional recombinant production methods, that is, a protein

that is not produced at all or is only produced at low levels. In another embodiment, the target protein is one that is readily produced using known expression systems, but for which there is a desire to achieve higher levels of expression.

[0077] Nucleic acids encoding a target protein may be obtained from any source using routine techniques well known in the art, including isolation from a genomic or cDNA library, amplification by PCR, or chemical synthesis.

[0078] The nucleotide sequence encoding a target protein used in the methods of the present invention comprises at the 5' end a linker DNA that is used for *in vivo* recombination with the linear vectors of the invention and further comprises at the 3' end a nucleotide sequence encoding a portion of the N-terminus of the reporter protein, including the N-terminal amino acids deleted in the linear vector and sufficient additional amino acids to allow *in vivo* recombination between the nucleotide sequence encoding a target protein and the linear vector when they are co-transformed into the host cell. In one embodiment, the sequence encoding a portion of the N-terminus of the reporter protein comprises at least 20 base pairs that overlap with the reporter protein-encoding sequence of the linear vector, *e.g.*, at least 30 or 40 base pairs. The addition of the 5' linker and the 3' reporter protein sequence to the nucleotide sequence encoding a target protein may be carried out using routine recombinant DNA techniques, *e.g.*, PCR and/or restriction enzyme cleavage and ligation.

[0079] The linker DNA of the invention must be of sufficient length and have sufficient sequence identity to a portion of the nucleotide sequence of the linear vector to allow *in vivo* recombination between the target protein-encoding nucleotide sequence and the linear vector when they are co-transformed into a host cell. In one embodiment, the linker DNA is more than 20 base pairs in length, *e.g.*, more than 30 or 40 base pairs in length. In a further embodiment, the linker DNA is at least 80% identical to the corresponding sequence on the linear vector, *e.g.*, at least 85%, 90%, 95%, or 99% identical.

[0080] In one embodiment, the linker DNA encodes a protease recognition sequence thereby allowing cleavage at the junction of the TFP and the target protein. For example, the linker DNA may encode a yeast kex2p- or Kex2-like protease recognition sequence (*e.g.*, an amino acid sequence comprising Lys-Arg, Arg-Arg, or Leu-Asp-Lys-Arg (SEQ ID NO:214)), a mammalian furin-recognition sequence (*e.g.*, an amino acid sequence comprising Arg-X-X-Arg), a factor Xa-recognition sequence (*e.g.*, an amino acid sequence comprising Ile-Glu-Gly-Arg (SEQ ID NO:215)), an enterokinase-recognition sequence (*e.g.*, an amino acid sequence comprising Asp-Asp-Lys), a subtilisin-recognition sequence (*e.g.*, an amino acid sequence comprising Ala-Ala-His-Tyr (SEQ ID NO:216)), a tobacco etch virus protease-recognition sequence (*e.g.*, an amino acid sequence comprising Glu-Asn-Leu-Tyr-Phe-Gln-Gly (SEQ ID NO:217)), a ubiquitin hydrolase-recognition sequence (*e.g.*, an amino acid sequence comprising Arg-Gly-Gly) or a thrombin-recognition sequence (*e.g.*, an amino acid sequence comprising Arg-Gly-Pro-Arg (SEQ ID NO:218)).

[0081] In another embodiment, the linker DNA encodes an affinity tag, *e.g.*, GST, MBP, NusA, thioredoxin, ubiquitin, FLAG, BAP, 6HIS, STREP, CBP, CBD, or S-tag.

[0082] In a further embodiment, the linker DNA encodes a restriction enzyme recognition site, *e.g.*, a SfiI site. In another embodiment, the linker DNA encodes a restriction enzyme recognition site and a protease recognition sequence (e.g., kex2p-like protease- or kex-2p-recognition sequence).

[0083] The present invention relates to a TFP identified by the methods of the invention or a derivative or fragment thereof. In one embodiment, the TFP is selected from the group consisting of TFP-9 (SEQ ID NO:29), TFP-13 (SEQ ID NO:31), TFP-17 (SEQ ID NO:33), TFP-18 (SEQ ID NO:35), TFP-19 (SEQ ID NO:37), TFP-20 (SEQ ID NO:39), TFP-21 (SEQ ID NO:41), TFP-25 (SEQ ID NO:43), TFP-27 (SEQ ID NO:45), TFP-11 (SEQ ID NO:61), TFP-22 (SEQ ID NO:63), TFP-29 (SEQ ID NO:65), TFP-34 (SEQ ID NO:67), TFP-38 (SEQ ID NO:69), TFP-39 (SEQ ID NO:129), TFP-43 (SEQ

ID NO:131), TFP-44 (SEQ ID NO:133), TFP-48 (SEQ ID NO:135), TFP-52 (SEQ ID NO:137), TFP-54 (SEQ ID NO:139), TFP-40 (SEQ ID NO:175), TFP-50 (SEQ ID NO:177), TFP-51 (SEQ ID NO:179), TFP-57 (SEQ ID NO:181), TFP-58 (SEQ ID NO:183), TFP-59 (SEQ ID NO:185), TFP-5 (SEQ ID NO:200), TFP-6 (SEQ ID NO:202), TFP-7 (SEQ ID NO:204), TFP-8 (SEQ ID NO:206), PpTFP-1 (SEQ ID NO:84), PpTFP-2 (SEQ ID NO:86), PpTFP-3 (SEQ ID NO:88), and PpTFP-4 (SEQ ID NO:90) or a derivative or fragment thereof.

[0084] The invention further relates to a TFP library comprising two or more of the TFPs identified by the methods of the invention or a fragment or derivative thereof. In one embodiment, the TFP library comprises TFPs identified as effective for a particular target protein. In another embodiment, the TFP library comprises TFPs identified as effective for more than one target protein. In a particular embodiment, the TFP library comprises two or more (*e.g.*, 4, 6, 8, 10, or 12 or more) TFPs selected from the group consisting of TFP-9 (SEQ ID NO:29), TFP-13 (SEQ ID NO:31), TFP-17 (SEQ ID NO:33), TFP-18 (SEQ ID NO:35), TFP-19 (SEQ ID NO:37), TFP-20 (SEQ ID NO:39), TFP-21 (SEQ ID NO:41), TFP-25 (SEQ ID NO:43), TFP-27 (SEQ ID NO:45), TFP-11 (SEQ ID NO:61), TFP-22 (SEQ ID NO:63), TFP-29 (SEQ ID NO:65), TFP-34 (SEQ ID NO:67), TFP-38 (SEQ ID NO:69), TFP-39 (SEQ ID NO:129), TFP-43 (SEQ ID NO:131), TFP-44 (SEQ ID NO:133), TFP-48 (SEQ ID NO:135), TFP-52 (SEQ ID NO:137), TFP-54 (SEQ ID NO:139), TFP-40 (SEQ ID NO:175), TFP-50 (SEQ ID NO:177), TFP-51 (SEQ ID NO:179), TFP-57 (SEQ ID NO:181), TFP-58 (SEQ ID NO:183), TFP-59 (SEQ ID NO:185), TFP-5 (SEQ ID NO:200), TFP-6 (SEQ ID NO:202), TFP-7 (SEQ ID NO:204), TFP-8 (SEQ ID NO:206), PpTFP-1 (SEQ ID NO:84), PpTFP-2 (SEQ ID NO:86), PpTFP-3 (SEQ ID NO:88), and PpTFP-4 (SEQ ID NO:90) or a derivative or fragment thereof.

[0085] In a further embodiment, the TFP library comprises six or more (*e.g.*, 8, 10, 12, or 15 or more) TFPs selected from the group consisting of TFP-9 (SEQ ID NO:29), TFP-13 (SEQ ID NO:31), TFP-17 (SEQ ID NO:33), TFP-

18 (SEQ ID NO:35), TFP-19 (SEQ ID NO:37), TFP-20 (SEQ ID NO:39), TFP-21 (SEQ ID NO:41), TFP-25 (SEQ ID NO:43), TFP-27 (SEQ ID NO:45), TFP-11 (SEQ ID NO:61), TFP-22 (SEQ ID NO:63), TFP-29 (SEQ ID NO:65), TFP-34 (SEQ ID NO:67), TFP-38 (SEQ ID NO:69), TFP-39 (SEQ ID NO:129), TFP-43 (SEQ ID NO:131), TFP-44 (SEQ ID NO:133), TFP-48 (SEQ ID NO:135), TFP-52 (SEQ ID NO:137), TFP-54 (SEQ ID NO:139), TFP-40 (SEQ ID NO:175), TFP-50 (SEQ ID NO:177), TFP-51 (SEQ ID NO:179), TFP-57 (SEQ ID NO:181), TFP-58 (SEQ ID NO:183), TFP-59 (SEQ ID NO:185), TFP-5 (SEQ ID NO:200), TFP-6 (SEQ ID NO:202), TFP-7 (SEQ ID NO:204), TFP-8 (SEQ ID NO:206), PpTFP-1 (SEQ ID NO:84), PpTFP-2 (SEQ ID NO:86), PpTFP-3 (SEQ ID NO:88), PpTFP-4 (SEQ ID NO:90), TFP-1 (SEQ ID NO:219), TFP-2 (SEQ ID NO:221), TFP-3 (SEQ ID NO:223), TFP-4 (SEQ ID NO:225), and TFP 32 (SEQ ID NO:208), or a derivative or fragment thereof.

[0086] The present invention further relates to a nucleic acid encoding a TFP identified by the methods of the invention or a derivative or fragment thereof. In one embodiment, the nucleic acid encodes a TFP selected from the group consisting of TFP-9 (SEQ ID NO:29), TFP-13 (SEQ ID NO:31), TFP-17 (SEQ ID NO:33), TFP-18 (SEQ ID NO:35), TFP-19 (SEQ ID NO:37), TFP-20 (SEQ ID NO:39), TFP-21 (SEQ ID NO:41), TFP-25 (SEQ ID NO:43), TFP-27 (SEQ ID NO:45), TFP-11 (SEQ ID NO:61), TFP-22 (SEQ ID NO:63), TFP-29 (SEQ ID NO:65), TFP-34 (SEQ ID NO:67), TFP-38 (SEQ ID NO:69), TFP-39 (SEQ ID NO:129), TFP-43 (SEQ ID NO:131), TFP-44 (SEQ ID NO:133), TFP-48 (SEQ ID NO:135), TFP-52 (SEQ ID NO:137), TFP-54 (SEQ ID NO:139), TFP-40 (SEQ ID NO:175), TFP-50 (SEQ ID NO:177), TFP-51 (SEQ ID NO:179), TFP-57 (SEQ ID NO:181), TFP-58 (SEQ ID NO:183), TFP-59 (SEQ ID NO:185), TFP-5 (SEQ ID NO:200), TFP-6 (SEQ ID NO:202), TFP-7 (SEQ ID NO:204), TFP-8 (SEQ ID NO:206), PpTFP-1 (SEQ ID NO:84), PpTFP-2 (SEQ ID NO:86), PpTFP-3 (SEQ ID NO:88), and PpTFP-4 (SEQ ID NO:90) or a derivative or fragment thereof. In one embodiment, the nucleic acid comprises a nucleotide sequence selected

from the group consisting of SEQ ID NOS:30, 32, 34, 36, 38, 40, 42, 44, 46, 62, 64, 66, 68, 70, 85, 87, 89, 91, 130, 132, 134, 136, 138, 140, 176, 178, 180, 182, 184, 186, 201, 203, 205, or 207 or a derivative or fragment thereof.

[0087] The invention further relates to a library of nucleic acids encoding two or more TFPs identified by the methods of the invention or a derivative or fragment thereof. In one embodiment, the library of nucleic acids encodes TFPs identified as effective for a particular target protein. In another embodiment, the library of nucleic acids encodes TFPs identified as effective for more than one target protein. In a particular embodiment, the library of nucleic acids encodes two or more (*e.g.*, 4, 6, 8, 10, or 12 or more) TFPs selected from the group consisting of TFP-9 (SEQ ID NO:29), TFP-13 (SEQ ID NO:31), TFP-17 (SEQ ID NO:33), TFP-18 (SEQ ID NO:35), TFP-19 (SEQ ID NO:37), TFP-20 (SEQ ID NO:39), TFP-21 (SEQ ID NO:41), TFP-25 (SEQ ID NO:43), TFP-27 (SEQ ID NO:45), TFP-11 (SEQ ID NO:61), TFP-22 (SEQ ID NO:63), TFP-29 (SEQ ID NO:65), TFP-34 (SEQ ID NO:67), TFP-38 (SEQ ID NO:69), TFP-39 (SEQ ID NO:129), TFP-43 (SEQ ID NO:131), TFP-44 (SEQ ID NO:133), TFP-48 (SEQ ID NO:135), TFP-52 (SEQ ID NO:137), TFP-54 (SEQ ID NO:139), TFP-40 (SEQ ID NO:175), TFP-50 (SEQ ID NO:177), TFP-51 (SEQ ID NO:179), TFP-57 (SEQ ID NO:181), TFP-58 (SEQ ID NO:183), TFP-59 (SEQ ID NO:185), TFP-5 (SEQ ID NO:200), TFP-6 (SEQ ID NO:202), TFP-7 (SEQ ID NO:204), TFP-8 (SEQ ID NO:206), PpTFP-1 (SEQ ID NO:84), PpTFP-2 (SEQ ID NO:86), PpTFP-3 (SEQ ID NO:88), and PpTFP-4 (SEQ ID NO:90) or a derivative or fragment thereof. In one embodiment, the library of nucleic acids comprises two or more (*e.g.*, 4, 6, 8, 10, or 12 or more) of the nucleotide sequences of SEQ ID NOS:30, 32, 34, 36, 38, 40, 42, 44, 46, 62, 64, 66, 68, 70, 85, 87, 89, 91, 130, 132, 134, 136, 138, 140, 176, 178, 180, 182, 184, 186, 201, 203, 205, or 207 or a derivative or fragment thereof.

[0088] In a further embodiment, the library of nucleic acids encodes six or more (*e.g.*, 8, 10, 12, or 15 or more) TFPs selected from the group consisting of TFP-9 (SEQ ID NO:29), TFP-13 (SEQ ID NO:31), TFP-17 (SEQ ID

NO:33), TFP-18 (SEQ ID NO:35), TFP-19 (SEQ ID NO:37), TFP-20 (SEQ ID NO:39), TFP-21 (SEQ ID NO:41), TFP-25 (SEQ ID NO:43), TFP-27 (SEQ ID NO:45), TFP-11 (SEQ ID NO:61), TFP-22 (SEQ ID NO:63), TFP-29 (SEQ ID NO:65), TFP-34 (SEQ ID NO:67), TFP-38 (SEQ ID NO:69), TFP-39 (SEQ ID NO:129), TFP-43 (SEQ ID NO:131), TFP-44 (SEQ ID NO:133), TFP-48 (SEQ ID NO:135), TFP-52 (SEQ ID NO:137), TFP-54 (SEQ ID NO:139), TFP-40 (SEQ ID NO:175), TFP-50 (SEQ ID NO:177), TFP-51 (SEQ ID NO:179), TFP-57 (SEQ ID NO:181), TFP-58 (SEQ ID NO:183), TFP-59 (SEQ ID NO:185), TFP-5 (SEQ ID NO:200), TFP-6 (SEQ ID NO:202), TFP-7 (SEQ ID NO:204), TFP-8 (SEQ ID NO:206), PpTFP-1 (SEQ ID NO:84), PpTFP-2 (SEQ ID NO:86), PpTFP-3 (SEQ ID NO:88), PpTFP-4 (SEQ ID NO:90), TFP-1 (SEQ ID NO:219), TFP-2 (SEQ ID NO:221), TFP-3 (SEQ ID NO:223), TFP-4 (SEQ ID NO:225), and TFP 32 (SEQ ID NO:208) or a derivative or fragment thereof. In one embodiment, the library of nucleic acids comprises six or more (*e.g.*, 8, 10, 12, or 15 or more) of the nucleotide sequences of SEQ ID NOS:30, 32, 34, 36, 38, 40, 42, 44, 46, 62, 64, 66, 68, 70, 85, 87, 89, 91, 130, 132, 134, 136, 138, 140, 176, 178, 180, 182, 184, 186, 201, 203, 205, 207, 209, 220, 222, 224, or 226 or a derivative or fragment thereof.

**[0089]** The term "fragment thereof," as applied to a TFP, refers to a polypeptide comprising of any portion of the amino acid sequence of the TFP, wherein the fragment substantially retains the ability to induce the secretion of a target protein to which it is fused.

**[0090]** The term "derivative thereof," as applied to a TFP, refers to a polypeptide consisting of an amino acid sequence that is at least 70% identical to the amino acid sequence of the TFP, wherein the polypeptide substantially retains the ability to induce the secretion of a target protein to which it is fused. In some embodiments, the derivative comprises an amino acid sequence that is at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identical to the amino acid sequence of the TFP. The derivative may comprise additions, deletions, substitutions, or a combination thereof to the amino acid

sequence of the TFP. Additions or substitutions also include the use of non-naturally occurring amino acids.

[0091] Preferably, any substitutions are conservative amino acid substitutions. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined within the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine).

[0092] The term "derivative thereof," as applied to a nucleic acid encoding a TFP, refers to a nucleic acid consisting of a nucleotide sequence that is at least 70% identical to the nucleotide sequence of the nucleic acid encoding the TFP, wherein the polypeptide encoded by the derivative substantially retains the ability to induce the secretion of a target protein to which it is fused. In some embodiments, the derivative comprises a nucleotide sequence that is at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identical to the nucleotide sequence of the nucleic acid encoding the TFP. The derivative may comprise additions, deletions, substitutions, or a combination thereof to the nucleotide sequence of the nucleic acid encoding the TFP.

[0093] Sequence identity is calculated by comparing two optimally aligned sequences over that region of comparison, determining the number of positions at which the identical amino acid residue or nucleotide occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the region of comparison (*i.e.*, the window size), and multiplying the result by 100 to yield the percentage of sequence identity. In one aspect, percent identity is calculated as the percentage of amino acid residues or nucleotides in the smaller

of two sequences which align with an identical amino acid residue or nucleotide in the sequence being compared, when four gaps in a length of 100 amino acids or nucleotides may be introduced to maximize alignment (Dayhoff, in Atlas of Protein Sequence and Structure, Vol. 5, p. 124, National Biochemical Research Foundation, Washington, D.C. (1972), incorporated herein by reference). A determination of identity is typically made by a computer homology program known in the art. An exemplary program is the Gap program (Wisconsin Sequence Analysis Package, Version 8 for UNIX, Genetics Computer Group, University Research Park, Madison, WI) using the default settings, which uses the algorithm of Smith and Waterman (Adv. Appl. Math., 1981, 2: 482-489, which in incorporated herein by reference in its entirety).

[0094] Examples of derivatives include, but are not limited to deletion mutants (*e.g.*, unidirectional deletion), addition of functional sequences (*e.g.*, glycosylation sites, restriction enzyme sites), and deletion or addition (*e.g.*, swapping) of pro-sequences or pre-sequences identified within TFPs. One of skill in the art can prepare derivatives of TFPs or nucleic acids encoding TFPs using routine mutagenesis techniques, such as those described in the references cited above, and identify derivatives that substantially retain the ability to induce the secretion of a target protein to which it is fused.

[0095] The term " substantially retains the ability to induce the secretion of a target protein to which it is fused," as used herein, refers to a fragment or derivative of a TFP which retains at least 50% of the ability of the parent TFP to induce secretion of a target protein to which it is fused. In some embodiments, at least 60, 65, 70, 75, 80, 85, 90, or 95% of the ability to induce the secretion of a target protein to which it is fused is retained. The ability to induce the secretion of a target protein may be determined by routine techniques well known in the art and described above.

[0096] One embodiment of the present invention relates to a library of nucleic acid fragments encoding TFPs, comprising 10 or more nucleic acid fragments (*e.g.*, 50, 100, 500, 100, or 2000 or more) identified by the methods of the

invention, wherein a library of pre-selected candidate TFPs was used in the screening.

[0097] Another embodiment of the present invention relates to a library of nucleic acid fragments encoding TFPs, comprising 10 or more nucleic acid fragments (*e.g.*, 50 or 100 or more) identified by the methods of the invention, wherein a library of pre-selected candidate TFPs obtained by transforming a plurality of reporter protein-deficient host cells with a variety of vectors comprising a library of nucleic acid fragments and a nucleotide sequence encoding a reporter protein, collecting cells that grow, isolating vectors from the cells, and isolating nucleic acid fragments from the vectors, thereby obtaining a TFP library comprising nucleic acid fragments which individually induce secretion of the reporter protein, was used in the screening.

[0098] A further embodiment of the present invention relates to a library of nucleic acid fragments encoding TFPs, comprising 10 or more nucleic acid fragments (*e.g.*, 50, 100, 500, or 1000 or more) identified by the methods of the invention, wherein a library of pre-selected candidate TFPs derived from sequences identified in a genome database by searching for (i) genes containing a pre-secretion signal homologous with those of one or more previously identified TFPs; (ii) genes comprising a secretion signal sequence, or (iii) genes encoding proteins passing through endoplasmic reticulum, was used in the screening.

[0099] A further embodiment of the present invention relates to a library of nucleic acid fragments encoding TFPs, comprising 10 or more nucleic acid fragments (*e.g.*, 50, 100, or 500 or more) identified by the methods of the invention, wherein a library of pre-selected candidate TFPs obtained by diversifying previously identified TFPs, was used in the screening.

[0100] The present invention further relates to a nucleic acid comprising a nucleotide sequence encoding a TFP identified by the methods of the invention and a nucleotide sequence encoding a target protein. In one embodiment, the TFP is selected from the group consisting of TFP-9, TFP-13, TFP-17, TFP-18, TFP-19, TFP-20, TFP-21, TFP-25, TFP-27, TFP-11, TFP-

22, TFP-29, TFP-34, TFP-38, TFP-39, TFP-43, TFP-44, TFP-48, TFP-52, TFP-54, TFP-40, TFP-50, TFP-51, TFP-57, TFP-58, TFP-59, TFP-5, TFP-6, TFP-7 and TFP-8 or a derivative or fragment thereof. In another embodiment, the target protein is selected from IL-2, IL-32, human growth hormone and human caspase-1 subunit P10. In a particular embodiment, the TFP is TFP-9, TFP-13, TFP-17, TFP-18, TFP-19, TFP-20, TFP-21, TFP-25, TFP-27, PpTFP-1, PpTFP-2, PpTFP-3, PpTFP-4 or a derivative or fragment thereof, and the target protein is IL-2. In another embodiment, the TFP is TFP-11, TFP-22, TFP-29, TFP-34 or TFP-38 or a derivative or fragment thereof, and the target protein is IL-32 alpha. In a further embodiment, the TFP is TFP-9, TFP-13, TFP-17, TFP-18, TFP-19, TFP-20, TFP-21, TFP-25, TFP-27, TFP-11, TFP-22, TFP-29, TFP-34 or TFP-38 or a derivative or fragment thereof, and the target protein is growth hormone.

[0101] The present invention further relates to methods of recombinantly producing a target protein using the TFPs of the invention. In one embodiment, the method comprises preparing a vector comprising a nucleotide sequence encoding a target protein operably linked to a nucleotide sequence encoding a TFP or a derivative or fragment thereof, transforming a host cell with the vector, and culturing the host cell under conditions in which the target protein is produced and secreted from the host cell. In one embodiment, the TFP is selected from the group consisting of TFP-9 (SEQ ID NO:29), TFP-13 (SEQ ID NO:31), TFP-17 (SEQ ID NO:33), TFP-18 (SEQ ID NO:35), TFP-19 (SEQ ID NO:37), TFP-20 (SEQ ID NO:39), TFP-21 (SEQ ID NO:41), TFP-25 (SEQ ID NO:43), TFP-27 (SEQ ID NO:45), TFP-11 (SEQ ID NO:61), TFP-22 (SEQ ID NO:63), TFP-29 (SEQ ID NO:65), TFP-34 (SEQ ID NO:67), TFP-38 (SEQ ID NO:69), TFP-39 (SEQ ID NO:129), TFP-43 (SEQ ID NO:131), TFP-44 (SEQ ID NO:133), TFP-48 (SEQ ID NO:135), TFP-52 (SEQ ID NO:137), TFP-54 (SEQ ID NO:139), TFP-40 (SEQ ID NO:175), TFP-50 (SEQ ID NO:177), TFP-51 (SEQ ID NO:179), TFP-57 (SEQ ID NO:181), TFP-58 (SEQ ID NO:183), TFP-59 (SEQ ID NO:185), TFP-5 (SEQ ID NO:200), TFP-6 (SEQ ID NO:202), TFP-

7 (SEQ ID NO:204), TFP-8 (SEQ ID NO:206), PpTFP-1 (SEQ ID NO:84), PpTFP-2 (SEQ ID NO:86), PpTFP-3 (SEQ ID NO:88), and PpTFP-4 (SEQ ID NO:90) or a derivative or fragment thereof. In a further embodiment, the target protein is selected from IL-2, IL-32, human growth hormone and human caspase-1 subunit P10.

[0102] The target protein may be recombinantly produced using any expression system known in the art. Preferably, the target protein is recombinantly expressed, *e.g.*, in bacterial, yeast, or mammalian cell cultures. Recombinant expression involves preparing a vector comprising a polynucleotide encoding the target protein, delivering the vector into a host cell, culturing the host cell under conditions in which the target protein is expressed, and separating the target protein. Methods and materials for preparing recombinant vectors and transforming host cells using the same, replicating the vectors in host cells and expressing biologically active foreign polypeptides and proteins are discussed above and described in Sambrook *et al.*, Molecular Cloning, 3rd edition, Cold Spring Harbor Laboratory, 2001 and Ausubel *et al.*, Current Protocols in Molecular Biology, John Wiley & Sons, New York 3rd edition, (2000), each incorporated herein by reference.

[0103] Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.* (MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989), and other laboratory manuals.

[0104] For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to

identify and select these integrants, a gene that encodes a selectable marker (*e.g.*, resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Various selectable markers include those that confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding the target protein or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

[0105] The target protein may be isolated from the medium in which the host cells are grown, by purification methods known in the art, *e.g.*, conventional chromatographic methods including immunoaffinity chromatography, receptor affinity chromatography, hydrophobic interaction chromatography, lectin affinity chromatography, size exclusion filtration, cation or anion exchange chromatography, high pressure liquid chromatography (HPLC), reverse phase HPLC, and the like. Still other methods of purification include those methods wherein the desired protein is expressed and purified as a fusion protein having a specific tag, label, or chelating moiety that is recognized by a specific binding partner or agent. The purified protein can be cleaved to yield the desired protein, or can be left as an intact fusion protein. Cleavage of the fusion component may produce a form of the desired protein having additional amino acid residues as a result of the cleavage process.

[0106] If the isolated target protein is not biologically active following the isolation procedure employed, various methods for "refolding" or converting the polypeptide to its tertiary structure and generating disulfide linkages, can be used to restore biological activity. Methods known to one of ordinary skill in the art include adjusting the pH of the solubilized polypeptide to a pH usually above 7 and in the presence of a particular concentration of a chaotrope. The selection of chaotrope is very similar to the choices used for inclusion body solubilization but usually at a lower concentration and is not necessarily the same chaotrope as used for the solubilization. It may be

required to employ a reducing agent or the reducing agent plus its oxidized form in a specific ratio, to generate a particular redox potential allowing for disulfide shuffling to occur in the formation of the protein's cysteine bridge(s). Some of the commonly used redox couples include cysteine/cystamine, glutathione (GSH)/dithiobis GSH, cupric chloride, dithiothreitol (DTT)/dithiane DTT, 2-mercaptoethanol (bME)/dithio-b(ME). To increase the efficiency of the refolding, it may be necessary to employ a cosolvent, such as glycerol, polyethylene glycol of various molecular weights, and arginine.

[0107] In one embodiment, the present invention relates to a linear vector comprising a nucleic acid fragment from a library of nucleic acid fragments and a nucleotide sequence encoding a N-terminal amino acid-deleted reporter protein. In another embodiment, the linear vector further comprises a target protein-encoding nucleotide sequence.

[0108] The present invention further relates to a plurality of reporter protein-deficient host cells transformed with a library of linear vectors of the invention. In one embodiment, the host cells are further transformed with a nucleic acid encoding a target protein.

[0109] The following examples are illustrative, but not limiting, of the method and compositions of the present invention. Other suitable modifications and adaptations of the variety of conditions and parameters normally encountered in clinical therapy and which are obvious to those skilled in the art are within the spirit and scope of the invention.

## EXAMPLE 1

## PREPARATION OF INVERTASE-DEFICIENT YEAST MUTANT

[0110] For rapid screening of the translational fusion partners (TFP) of non-producible proteins, an automatic screening system was established through

the evaluation of cell growth in a sucrose medium using yeast invertase as a reporter.

[0111] A yeast strain having no invertase activity was required to use an invertase gene as a reporter for the positive screening of useful TFP. Thus, the chromosomal *SUC2* gene of wild type yeast was deleted. In order to prepare a *SUC2* deletion cassette, a plasmid pRB58 (Carlson *et al.*, *Cell 20*:145 (1982)) was digested with *Eco*RI and *Xho*I, and a *SUC2* coding gene was recovered and introduced into *Eco*RI-*Xho*I sites of pBluescript KS+ (Stratagene, USA), thus generating pBIΔBX. As shown in FIG. 1, an *URA3* gene having a repeat sequence of 190 bp (Tc190) (Bae *et al.*, *Yeast 21*:437 (2004)) at both ends was inserted into *Hin*dIII-*Xba*I sites of the *SUC2* gene contained in pBIΔBX, thus generating pBIU. The pBIU was digested with *Eco*RI and *Xho*I, and was transformed into *S. cerevisiae* Y2805 (*Mat a ura3 SUC2 pep4::HIS3 GAL1 can1*) and Y2805Δ*gal1* (*Mat a ura3 SUC2 pep4::HIS3 gal1 can1*) strain (SK Rhee, Korea Research Institute of Bioscience and Biotechnology) according to a lithium acetate method (Hill *et al.*, *Nucleic Acids Res. 19*:5791 (1991)). The transformants, Y2805Δ*suc2U* (*Mat a suc2::URA3 pep4::HIS3 GAL1 can1)*, Y2805Δ*gal1*Δ*suc2U* (*Mat a suc2::URA3 pep4::HIS3 gal1 can1)*, were selected in a selection medium lacking uracil.

[0112] To evaluate the invertase activity of the transformed cells, a single colony was cultured in two media containing glucose and sucrose, respectively, as the sole carbon source. As a result, the colonies grew normally in the glucose medium, but grew very slowly in the sucrose medium compared to a control. In order to investigate the amount of invertase secreted into the culture medium, the *SUC2*+ strain and the Δ*suc2* strain were cultured on YPD media (1% yeast extract, 2% Bacto-peptone and 2% glucose). Proteins contained in the culture supernatants were separated by SDS-PAGE, and the gel was incubated in a sucrose solution for 30 min and subjected to zymogram analysis using a dye, TTC (2, 3, 5-triphenyl-tetrazolium chloride). As shown in FIG. 2, the Δ*suc2* strain was found to lose most of its invertase activity. However, the mutant strain had a problem of growing even at very

slow rates in the sucrose medium. This is believed to be because cells partially grow by gluconeogenesis through the function of mitochondria. Thus, to solve this problem, antimycin A, an inhibitor of mitochondrial electron transport, was added to the medium to block cell growth. As a result, the growth of the mutant strain was completely inhibited in the YPSA (1% yeast extract, 2% Bacto-peptone, 2% sucrose, 1 μg/ml antimycin A, and 2% agar) or YPSGA (1% yeast extract, 2% Bacto-peptone, 2% sucrose, 1 μg/ml antimycin A, and 2% agar) medium containing antimycin A (FIG. 3).

[0113] In order to recover uracil auxotrophy of the selected strain, Y2805Δ*suc2U* (*Mat a suc2::URA3 pep4::HIS3 GAL1 can1)* and Y2805Δ*gal1*Δ*suc2U* (*Mat a suc2::URA3 pep4::HIS3 gal1 can1)*, with a *URA3* vector containing a TFP library, it was necessary to remove the *URA3* gene which was used for the deletion of the *SUC2* gene. To do this, cells were cultured in a medium containing 5-fluoroorotic acid (5-FOA) and selected for loss of the *URA3* gene, thus obtaining *URA3* pop-out strains, Y2805Δ*suc2* (*Mat a ura3 suc2::Tc190 pep4::HIS3 GAL1 can1)* and Y2805Δ*gal1*Δ*suc2* (*Mat a ura3 suc2::Tc190 pep4::HIS3 gal1 can1)* (FIG. 1). Southern blotting was carried out to confirm the deletion of the *SUC2* gene on the chromosome, as expected, and the *URA3* gene was deleted (popped-out) from the integration locus (FIG. 4). When chromosomal DNA from *S. cerevisiae* Y2805 was treated with *Eco*RI and analyzed by Southern blotting using a *SUC2* gene as a probe, a fragment of about 4.3 kb was detected. This size increased to about 5.0 kb when a *URA3* gene was inserted (Y2805Δ*gal1*Δ*suc2U*), and decreased to about 3.7 kb when the *URA3* gene was popped-out (Y2805Δ*gal1*Δ*suc2*). As shown in FIG. 4, as expected, the *SUC2* gene was obviously deleted, and the *URA3* gene was lost (popped-out).

## EXAMPLE 2

## DEVELOPMENT OF AUTOMATIC SCREENING SYSTEM USING AN INVERTASE AS A SECRETION REPORTER

[0114] The invertase deficient strain was evaluated for the possibility of being automatically screened in a sucrose medium through the expression of a protein fused to invertase, using two human therapeutic proteins, a human serum albumin (HSA) which is well secreted in yeast, and a human interleukin-2 (IL-2) which is hardly secretable in yeast.

[0115] Three plasmids, pYGAP-SNS-SUC2, pYGAP-HSA-SUC2 and pYGAP-hIL2-SUC2, were constructed to test for automatic selection on sucrose media. For the construction of pYGAP-SUC2 containing an invertase gene (*SUC2*, YIL162W) expression cassette under the control of the yeast *GAPDH* promoter, pST-SUC2 was constructed first by subcloning a PCR product containing *SUC2* gene amplified from pBIΔBX (FIG. 1) using primers SUC-F (SEQ ID NO. 1) and SUC-R (SEQ ID NO. 2) into pST-Blue-1 (Novagen, USA). PCR was carried out with *Pfu* polymerase (Stratagene, USA) or Ex-Taq DNA polymerase (TaKaRa Korea Biomedical Inc., Seoul, Korea). PCR conditions included one cycle of 94°C for 5 min, and 25 cycles of 94°C for 30 sec, 55°C for 30 sec and 72°C for 2 min, followed by one final cycle of 72°C for 7 min. The *Eco*RI-*Sal*I fragment containing *SUC2* from pST-SUC2 was subcloned into *Eco*RI-*Sal*I digested YGAPα-HIR harboring a *GAPDH* promoter instead of the *GAL10* promoter of YEGα-HIR525 (Sohn *et al.*, *Process Biochem. 30*:653 (1995)), and the resulting plasmid was named pYGAP-SUC2. To facilitate the fusion of foreign genes with *SUC2* and induce *in vivo* cleavage of the fused proteins by yeast dipeptidyl protease Kex2p (Mizuno K *et al.*, *Biochem. Biophys. Res. Commun. 156*:246 (1988)) during secretion, an artificial sequence for two *Sfi*I and a *Not*I recognition sites and a sequence coding for Kex2p cleavage site (Leu-Asp-Lys-Arg (SEQ ID NO:214)) were in-frame added between a secretion signal sequence (19 amino

acids) and a *SUC2* mature sequence (513 amino acids) of *SUC2* by PCR. Two PCR fragments, PCR-A containing a *GAPDH* promoter and a *SUC2* secretion signal sequence amplified using primers GAP-F (SEQ ID NO:3) and SUCSS-R (SEQ ID NO:4) and PCR-B containing a mature part of *SUC2* amplified from pYGAP-SUC2 using primers SUCM-F (SEQ ID NO:5) and SUC-R (SEQ ID NO:2) were amplified from pYGAP-SUC2, respectively. Both fragments were subcloned into pST-Blue-1 and recovered by *Sac*I-*Not*I digestion for PCR-A and *Not*I-*Sal*I digestion for PCR-B. Enzyme digested PCR-A and PCR-B were co-ligated into *Sac*I-*Sal*I digested pYGAP-SUC2 and the resulting plasmid was named pYGAP-SNS-SUC2. For the construction of a plasmid, pYGAP-HSA-SUC2, containing an in-frame fused gene between human serum albumin (HSA) with *SUC2*, the HSA gene was amplified from pYHSA5 (Kang *et al.*, J. *Microbiol. Biotechnol.* *8*:42 (1998)) using primers HSA-F (SEQ ID NO:6) and HSA-R (SEQ ID NO:7) and subcloned in pST-Blue-1. A *Sfi*I digested DNA containing the *HSA* gene was subcloned into the *Sfi*I digested pYGAP-SNS-SUC2 vector. The resulting plasmid was named pYGAP-HSA-SUC2. For the construction of a plasmid, pYGAP-hIL2-SUC2, containing an in-frame fused gene between human interleukin-2 (hIL2) with *SUC2*, the hIL2 gene was amplified from pT7-hIL2 (JK Jung, Korea Research Institute of Bioscience and Biotechnology) using primers IL2-F (SEQ ID NO:8) and IL2-R (SEQ ID NO:9) and subcloned into pST-Blue-1. Then a plasmid pYGAP-hIL2-SUC2 was constructed by the subcloning of a *Sfi*I digested hIL2 fragment into the *Sfi*I digested pYGAP-SNS-SUC2 vector.

**[0116]** The pYGAP-HSA-SUC2 vector expressing a fusion protein of human serum albumin and invertase, the pYGAP-hIL2-SUC2 expressing a fusion protein of IL-2 and invertase, and the pYGAP-SNS-SUC2 expressing only invertase were individually transformed into a yeast strain (Y2805Δ*suc2*) which is deleted for its endogenous invertase gene and thus unable to grow in a sucrose medium. The transformed cells were spread onto a UD plate (0.67% yeast nitrogen base without amino acids, 0.77 g/l amino acid mixture, 2% glucose and 2% agar) containing glucose as a sole carbon source and YPSA

media (1% yeast extract, 2% Bacto-peptone, 2% sucrose, 1 μg/ml antimycin A, and 2% agar) containing sucrose as a sole carbon source and cell growth of each transformation was observed (FIG. 5). When cells were transformed with pYGAP-SNS-SUC2 expressing invertase, they normally grew in both carbon sources. Similarly, when cells were transformed with pYGAP-HSA-SUC2 having a fusion of HSA at the N-terminus of invertase, they grew well using both carbon sources. In contrast, when cells were transformed with pYGAP-hIL2-SUC2 having a fusion of IL2 instead of HSA, they grew normally on the glucose medium but hardly grew on the sucrose medium. This inability of the pYGAP-hIL2-SUC2-transformed cells to grow in the sucrose medium was believed to be caused by the IL-2 being unable to be secreted from the cells and leading to a block of the secretion of invertase fused thereto. These results suggested a positive selection system using an invertase as a reporter for a secretion signals and a fusion partner (a translational fusion partner, TFP) from any sources of DNA enhancing the secretion of non- or hardly-secretable proteins such as human IL2.

## EXAMPLE 3

## PREPARATION OF VECTORS FOR THE CONSTRUCTION OF TRANSLATIONAL FUSION PARTNER (TFP) LIBRARY

[0117] Several vectors were designed for the construction of a TFP library from genomic DNA or a cDNA library from any source. For the construction of a TFP library from cDNA, a plasmid YGaINV was constructed (FIG. 6). A PCR was carried out to amplify a DNA fragment encoding invertase from pYGAP-hIL2-SUC2 using two PCR primers, SfiI-SUC-F (SEQ ID NO:10) and SUC-Xho-R (SEQ ID NO:11). PCR conditions included one cycle of 94°C for 5 min, and 25 cycles of 94°C for 30 sec, 55°C for 30 sec and 72°C for 2 min, followed by one final cycle of 72°C for 7 min. Then an *Eco*RI-*Sal*I digested PCR fragment was ligated to *Eco*RI-*Sal*I digested YEGα-HIR525

and the resulting plasmid was named YGaINV (FIG. 6). For the construction of a TFP library from partially digested genomic DNA, three vectors, YGaF0INV, YGaF1INV and YGaF2INV, each containing one of three different reading frames of the *SUC2* gene were constructed (FIG. 7). Three different PCR amplifications were performed from YGaINV as a template using a common forward primer Gal100-F (SEQ ID NO:12) and three reverse primers with different reading frames, Xho-F0-R (SEQ ID NO:13), Xho-F1-R (SEQ ID NO:14), and Xho-F2-R (SEQ ID NO:15). PCR was done using a *Pfu* polymerase (Stratagene, USA). PCR conditions included one cycle of 94°C for 5 min, and 25 cycles of 94°C for 30 sec, 55°C for 30 sec and 72°C for 2 min, followed by one final cycle of 72°C for 7 min. Three PCR fragments were eluted from agarose gel and digested with *Sfi*I. Then they were subcloned into *Sfi*I digested YGaINV, respectively. Three resulting plasmids were named YGaF0INV, YGaF1INV and YGaF2INV (FIG. 7).

## EXAMPLE 4

## CONSTRUCTION OF CDNA LIBRARY FUSED TO YEAST INVERTASE

[0118] For the construction of a cDNA library, total RNA was isolated from yeast *S. cerevisiae* Y2805 (*Mat a ura3 his3 pep4::HIS3 can1*). Yeast cells were cultivated to mid-exponential phase in YPD media (2% yeast extract, 1% Bacto-peptone and 2% glucose). Total RNA was isolated by a method described in Elion *et al.* (Elion *et al.*, *Cell 39*:663 (1984)). Purification of poly(A)$^+$ mRNA from the total RNA was carried out using an Oligotex mRNA kit (Qiagen, Germany). cDNA was synthesized from the isolated mRNA using a SMART cDNA synthesis kit (BD Bioscience, USA). A specially designed primer ASA24N6 (SEQ ID NO:16) was used for the synthesis of the first strand cDNA instead of a primer included in the SMART kit. Because the primer ASA24N6 was designed to contain a *Sfi*I recognition site and a random hexameric sequence, it was used for the synthesis of the first strand

cDNA from mRNA by reverse transcription as in the method described in the instruction manual of SMART kit (FIG. 8). Primer ASA24N6 could randomly bind to any position of mRNA due to its random hexameric sequence. Thus, most of the first stranded cDNA amplified by using this method contained the 5' partial sequence encoding the N-terminal part of yeast genes. The first stranded cDNA library with 5' partial sequences was used as a PCR template for double stranded cDNA synthesis with the 5' PCR primer of SMART kit (BD Bioscience, USA) and the primer ASA24 (SEQ ID NO:17). The resulting PCR products contained numerous 5' partial fragments of cDNA with *Sfi*I sites at both ends. PCR conditions included one cycle of 95°C for 20 sec, and 20 cycles of 95°C for 30 sec, 68°C for 6 min as recommended in the kit. Amplified cDNA was treated with phenol/chloroform/isoamyl alcohol (25:24:1) and precipitated with 2 volumes of ethanol and 0.1 volume of 3 M sodium acetate (pH 5.0). Recovered cDNA was digested with *Sfi*I at 50°C for 2 hours and then fractionated using agarose gel electrophoresis. 0.5 to 1 kb DNA was isolated from the gel using a gel extraction kit (Bioneer, Korea). Extracted DNA was ligated into *Sfi*I digested YGaINV vector (FIG. 6) and transformed into *E. coli* DH5α. Transformed *E. coli* was plated on LB media containing ampicillin (1% Bacto-peptone, 0.5% yeast extract, 1% NaCl with 50 μg/ml ampicillin) and incubated at 37°C overnight. About $5\times10^4$ *E. coli* colonies were pooled with sterile distilled water and the total plasmids containing random primed cDNA library fused to the *SUC2* gene were isolated by using a plasmid isolation kit (Bioneer, Korea).

## EXAMPLE 5

## CONSTRUCTION OF GENOMIC DNA LIBRARY FUSED TO YEAST INVERTASE

**[0119]** The TFP library constructed in Example 4 was obtained from a cDNA library which were synthesized from a pool of mRNA. Because the mRNA of

a highly expressed gene is usually abundant compared to that of a poorly expressed gene, a TFP library could be biased with those from highly expressed genes. Furthermore, some genes are completely repressed at a point of the growth phase and thus, they could not be amplified in a TFP library even though they were good candidates for a TFP. To solve such problems, genomic DNA was also used for the construction of a TFP library. As shown in FIG. 9, genomic DNA of *S. cerevisiae* Y2805 was partially digested with *Sau*3AI and incubated at 70°C for 10 min to inactivate the enzyme. The DNA was 2 bases filled with Klenow fragment and 0.2 mM of dTTP and dCTP at 25° for 1 hour and then 0.5 to 1 kb DNA was isolated from an agarose gel. In addition, vectors YGaF0INV, YGaF1INV and YGaF2INV (FIG. 7) were digested with *Xho*I. After inactivation of the enzyme at 70°C for 10 min, the vectors were also 2 bases filled with Klenow fragment and 0.2 mM of dTTP and dCTP and purified from an agarose gel. Each vector was ligated with the partially digested genomic DNA and transformed into *E. coli* DH5α, respectively. Transformed *E. coli* was plated on LB media containing ampicillin (1% Bacto-peptone, 0.5% yeast extract, 1% NaCl with 50 μg/ml ampicillin) and incubated at 37°C overnight. About $2\times10^5$ *E. coli* colonies obtained from three different vectors were pooled with sterile distilled water and the total plasmids containing genomic DNA library fused to the *SUC2* gene were isolated by using a plasmid isolation kit (Bioneer, Korea).

## EXAMPLE 6

## CONSTRUCTION OF TFP LIBRARY SECRETING INVERTASE

[0120] For the first selection of a TFP library secreting invertase from the genomic and cDNA libraries constructed in Example 4 and 5, library DNA was transformed into *S. cerevisiae* Y2805Δ*gal1*Δ*suc2*(*Mat a ura3 suc2::Tc190 pep4::HIS3 gal1 can1*) according to a lithium acetate method (Hill *et al.*, *Nucleic Acids Res. 19*:5791 (1991)). Y2805Δ*gal1*Δ*suc2* cannot

use sucrose and galactose as carbon sources due to the deletion of both genes. Transformed cells were spread on both UD media (0.67% yeast nitrogen base without amino acids, 0.77 g/l amino acid mixture, 2% glucose and 2% agar) and YPSGA media (1% yeast extract, 2% Bacto-peptone, 2% sucrose, 0.3% galactose, 1 μg/ml antimycin A, and 2% agar) and incubated at 30°C for 4 to 6 days. Around 3,000 and 1,000 transformants were obtained from the cDNA and genomic DNA library, respectively. All transformants grown on YPSGA media were transferred to a UD plate with a toothpick and incubated at 30°C for 2 days. Total DNA was isolated from the pooled cells using glass beads and then the DNA was precipitated with ethanol. To recover the plasmid containing TFP library, total DNA was retransformed into *E. coli* DH5α. Transformed *E. coli* was plated on LB media containing ampicillin (1% Bacto-peptone, 0.5% yeast extract, 1% NaCl with 50 μg/ml ampicillin) and incubated at 37°C overnight. Around $2\times10^4$ *E. coli* transformants were obtained and collected with sterile distilled water for the isolation of total plasmids using a plasmid isolation kit (Bioneer, Korea). Thus, a TFP pool containing up to 4,000 TFPs which individually induce the secretion of invertase was constructed. Nucleotide sequencing of the randomly selected plasmids from the library revealed that all of the TFPs were originated from yeast genes individually encoding different secretory proteins.

## EXAMPLE 7

## CONSTRUCTION OF TFP LIBRARY VECTOR APPLICABLE TO MANY TARGET PROTEINS THROUGH *IN VIVO* RECOMBINATION

[0121] Around 4,000 TFPs having a potential to secrete invertase were collected in Example 6. For the development of a TFP library vector which can be easily applicable to any target gene, a simple *in vivo* recombination system was designed. A vector, YGadV45 (FIG. 10), was first constructed for the in-frame insertion of any target protein gene between the TFP library and

the *SUC2* gene through *in vivo* recombination. YGadV45 contains a defective *SUC2* (d*SUC2*) which is an N-terminal 45 amino acid deleted *SUC2* and thus, having no invertase activity. The vector was also designed to contain a *Not*I and two *Sfi*I recognition sequences, a linker sequence as a recombination target and a *Swa*I recognition sequence in front of the d*SUC2* for the simple insertion of a TFP library and target gene through *in vivo* recombination. A PCR was carried out from a template YGaINV using a forward primer INV45-F (SEQ ID NO:18) and a reverse primer SUC-Xho-R (SEQ ID NO:11) and *Pfu* polymerase (Stratagene, USA). PCR conditions included one cycle of 94°C for 3 min, and 25 cycles of 94°C for 30 sec, 55°C for 30 sec and 72°C for 90 sec, followed by one final cycle of 72°C for 7 min. From the PCR, an N-terminal modified defective *SUC2* gene fragment was obtained. A *Not*I-*Sal*I digested PCR fragment was subcloned into a *Not*I-*Sal*I digested vector YGaINV (FIG. 6) and the resulting plasmid was named YGadV45 (FIG. 10). For the construction of a TFP library in YGadV45, the TFP library obtained in Example 6 was digested with *Sfi*I and fractionated in an agarose gel. Around 0.5 to 1 kb DNA fragments were isolated from the gel using a gel extraction kit (Bioneer, Korea). Purified DNA was subcloned into *Sfi*I digested YGadV45 (FIG. 10) and transformed into *E. coli* DH5α. Transformed *E. coli* was plated on LB media with ampicillin (1% Bacto-peptone, 0.5% yeast extract, 1% NaCl with 50 μg/ml ampicillin) and incubated at 37°C overnight. Around $5\times10^4$ *E. coli* transformants were collected with sterile distilled water for the isolation of total plasmids. Total plasmids were isolated using a plasmid isolation kit (Bioneer, Korea). The isolated vectors contained TFPs selected in Example 6 fused to a defective *SUC2*. Thus, transformation of this TFP library vector into *S. cerevisiae* Y2805Δ*gal1*Δ*suc2* (*Mat a ura3 suc2::Tc190 pep4::HIS3 gal1 can1*) gave thousands of transformants on a UD plate (0.67% yeast nitrogen base without amino acids, 0.77 g/l amino acid mixture, 2% glucose and 2% agar) but no transformants on YPSGA media (1% yeast extract, 2% Bacto-peptone, 2% sucrose, 0.3% galactose, 1 μg/ml antimycin A, and 2% agar). Thus, it could greatly reduce the background

level of selection on YPSGA media. Only the cells with a vector harboring an in-frame inserted target gene between TFP and *SUC2* could grow on YPSGA after correct *in vivo* recombination. The TFP library vectors contained a rare cutting restriction enzyme *Swa*I site and a linker sequence between the TFP library and the d*SUC2* for linearization and for homologous recombination, respectively.

## EXAMPLE 8

## AUTOSELECTION OF AN OPTIMAL TFP SECRETING A TARGET PROTEIN FROM TFP LIBRARY

**[0122]** For the in-frame fusion of target proteins through *in vivo* recombination to the TFP library vectors developed in Example 7, a target gene must have a linker DNA in the 5'-end and a N-terminal part of *SUC2* in the 3'-end. To add such a sequence to the end of a target gene, overlap extension PCR was used. A first step PCR was carried out for the amplification of a target gene encoding a mature protein using a target specific forward primer KR-target-F (SEQ ID NO:19) and a target specific reverse primer Target-INV-R (SEQ ID NO:20) from a plasmid containing target gene. Separately, another PCR for the amplification of a N-terminal part of *SUC2* which will be fused to the 3'-end of a target gene was also carried out using a forward primer KR-Inv-F (SEQ ID NO:21) and a reverse primer Inv500-R (SEQ ID NO:22) from YGaINV (FIG. 6). PCR was performed with *Pfu* polymerase (Stratagene, USA) and PCR conditions included one cycle of 94°C for 3 min, and 25 cycles of 94°C for 30 sec, 55°C for 30 sec and 72°C for 90 sec, followed by one final cycle of 72°C for 7 min. Then a second PCR was done from the two DNA fragments amplified in the first step using a forward primer LNK40 (SEQ ID NO:23) and a reverse primer Inv500-R (SEQ ID NO:22). The resulting fragment (insert fragment) harbored 40 nucleotides of linker DNA in the 5'-end and 500 bp of DNA encoding the Kex2p

recognition site (Leu-Asp-Lys-Arg (SEQ ID NO:214)) and a N-terminal part of invertase in the 3'-end, respectively. For *in vivo* recombination, the insert fragment was mixed at a 2:1 ratio with *SwaI* digested TFP library vectors constructed in Example 7 and used for transformation into *S. cerevisiae* Y2805Δ*gal1*Δ*suc2* (*Mat a ura3 suc2::Tc190 pep4::HIS3 gal1 can1*) (FIG. 11). Transformed cells were spread on YPSGA media (1% yeast extract, 2% Bacto-peptone, 2% sucrose, 0.3% galactose, 1 μg/ml antimycin A, and 2% agar) and incubated for 5 days. Only an in-frame fusion of the insert fragment with a vector containing a proper TFP through *in vivo* recombination could support cell growth on YPSGA media. Thus, using this method, an optimal TFP for any target protein could be retrieved by simple selection of growing cells on YPSGA media.

## EXAMPLE 9

## AUTOSELECTION OF AN OPTIMAL TFP USING A DUAL REPORTER SYSTEM WITH LIPASE AND INVERTASE

**[0123]** An autoselection system using a reporter, invertase, as described in Example 8 was very useful for the screening of an optimal TFP for a target protein such as IL2 which blocks the secretion of invertase completely as found in Example 2. As dozens of colonies could grow on sucrose media, it was easy to select an optimal TFP from the TFP library. Fusion of some target proteins, however, could not completely block the secretion of invertase even though a weak TFP was connected. Such leaky colonies could also grow on sucrose media. Thus, considerable number of colonies should be tested for their secretion level to select an optimal TFP. To solve such a time-consuming problem, a simple selection method was developed to identify a colony having a high protein secretion level with a halo-forming reporter, lipase, on a tributyrin-containing plate. A gene encoding lipase (CalB, lipase B of *Candida antarctica*) was in-frame fused to the 5'end of invertase. Using

this dual reporter system, transformants could be selected with both invertase and lipase activity on YPSGA media containing tributyrin, simultaneously. Colonies secreting protein at a high level could be simply determined with the size of halo formed around the colonies. As shown in FIG. 12, construction of a dual reporter was done by three steps of PCR. A 1 kb PCR fragment containing CalB was first amplified using a CalB forward primer KR-CalB-F (SEQ ID NO:24) and a reverse primer CalB-Inv-R (SEQ ID NO:25) from a plasmid pLGK-Lip14* containing a mutant CalB gene (SY Kim, Ph.D. thesis, Yonsei University, Korea, 2001). Separately, a 0.5 kb PCR fragment containing a 5' partial *SUC2* gene was amplified from YGaINV (FIG. 6) using a forward primer KR-Inv-F (SEQ ID NO:21) and a reverse primer Inv500-R (SEQ ID NO:22). PCR was performed with *Pfu* polymerase (Stratagene, USA) and PCR conditions included one cycle of 94°C for 3 min, and 25 cycles of 94°C for 30 sec, 55°C for 30 sec and 72°C for 90 sec, followed by one final cycle of 72°C for 7 min. Then, a second PCR was done from the two DNA fragments amplified in the first step using a forward primer KR-CalB-F (SEQ ID NO:24) and a reverse primer Inv500-R (SEQ ID NO:22). Separately, a PCR for a target gene was amplified using primers KR-Target-F (SEQ ID NO:19) and Target-CalB-R (SEQ ID NO:26) from a plasmid containing a target gene as described in Example 8. The third PCR was done using a forward primer LNK40 (SEQ ID NO:23) and a reverse primer Inv500-R (SEQ ID NO:22) from a template mixture of a target gene and CalB fused with a partial *SUC2* gene. The resulting DNA fragment (insert fragment) consisted of 40 nucleotides of linker, a target gene, Kex2p cleavage site (Leu-Asp-Lys-Arg (SEQ ID NO:214)), CalB, Kex2p cleavage site (Leu-Asp-Lys-Arg (SEQ ID NO:214)) and 500 bp of 5' partial SUC2 gene in order. For *in vivo* recombination, PCR amplified insert fragment was mixed at a 2:1 ratio with *Swa*I digested TFP library vectors constructed in Example 7 and used for transformation into *S. cerevisiae* Y2805Δ*gal1*Δ*suc2* (*Mat a ura3 suc2::Tc190 pep4::HIS3 gal1 can1*). Transformed cells were spread on YPSGA (1% yeast extract, 2% Bacto-peptone, 2% sucrose, 0.3% galactose, 1 μg/ml antimycin A,

and 2% agar) for selection with invertase activity and YPSGAT (1% yeast extract, 2% Bacto-peptone, 2% sucrose, 0.3% galactose, 1 μg/ml antimycin A, 1% tributyrin, and 2% agar) for selection with invertase and lipase activities, respectively. The transforming plates were incubated at 30°C for 5 days. Colonies secreting a target protein, lipase and invertase were formed on both YPSGA and YPSGAT plates. As expected, different sizes of halo were formed around colonies. The size of the halo was comparatively proportional to the secreted lipase activity. Thus, it was easy to select a colony with high secretion level directly from the transforming plate (FIG. 13).

## EXAMPLE 10

## NOVEL TFP SELECTED FROM TFP LIBRARY FOR THE SECRETION OF HUMAN INTERLEUKIN-2

[0124] As an example for identifying optimal TFPs using a method developed in this invention, a hardly secretable protein, human interleukin-2 (hIL2) was tried. An insert fragment containing the human IL2 gene and a 500 bp N-terminal part of *SUC2* was amplified using PCR as described in Example 8 (FIG. 11). A PCR was carried out using a forward primer KR-IL2-F (SEQ ID NO:27) and a reverse primer IL2-INV-R (SEQ ID NO:28) from pT7-hIL-2 (JK Jung, Korea Research Institute of Bioscience and Biotechnology) as a template. Separately, another PCR for the amplification of an N-terminal part of *SUC2* to be fused to the 3'-end of the IL2 gene was also carried out using a forward primer KR-Inv-F (SEQ ID NO:21) and a reverse primer Inv500-R (SEQ ID NO:22) from YGaINV (FIG. 6). Then the second PCR was done from the two DNA fragments amplified in the first step using a forward primer LNK40 (SEQ ID NO:23) and Inv500-R (SEQ ID NO:22). The resulting fragment (insert fragment) harbored a 40 nucleotide linker DNA containing a Kex2p recognition sequence (Leu-Asp-Lys-Arg (SEQ ID NO:214)), IL2, an additional Kex2p recognition sequence, and an N-terminal part of invertase in

order. This fragment was co-transformed with the *Swa*I digested TFP library vector constructed in Example 7 into *S. cerevisiae* Y2805Δ*gal1*Δ*suc2* (*Mat a ura3 suc2::Tc190 pep4::HIS3 gal1 can1*). Transformed cells were spread on both UD media (0.67% yeast nitrogen base without amino acids, 0.77 g/l amino acid mixture, 2% glucose and 2% agar) and YPSGA media (1% yeast extract, 2% Bacto-peptone, 2% sucrose, 0.3% galactose, 1 μg/ml antimycin A, and 2% agar) and incubated at 30°C for 5 days. Around $2\times10^4$ transformants were obtained in UD plates but about 100 transformants were obtained in YPSGA. Thirty randomly selected transformants growing on YPSGA was cultivated on YPD broth. Total DNA was isolated and retransformed into *E. coli* DH5α. Transformed *E. coli* was plated on LB media with ampicillin (1% Bacto-peptone, 0.5% yeast extract, 1% NaCl with 50 μg/ml ampicillin) and incubated at 37°C overnight. Plasmids were isolated from each *E. coli* transformant using a plasmid extraction kit (Bioneer, Korea). To analyze the sequence of each TFP, a sequencing primer GAL100-F (SEQ ID NO:12) binding to the *GAL10* promoter was used for all plasmids containing TFPs. Nucleotide sequences were determined by Genotech Co. (Taejon, Korea) using an automated sequencing unit (ABI Prism 377; PE Biosystems, Foster City, CA, USA). The sequences were analyzed by a BLAST search of the *Saccharomyces* Genome Database (www.yeastgenome.org). As a result, nine novel TFPs and a known TFP (TFP-3) (WO 2005/068658) were identified from plasmids isolated from 30 colonies which grew on YPSGA media. The isolated plasmids were named pYHTS-TFP9, pYHTS-TFP13, pYHTS-TFP17, pYHTS-TFP18, pYHTS-TFP19, pYHTS-TFP20, pYHTS-TFP21, pYHTS-TFP25, and pYHTS-TFP27, respectively. The nine novel TFPs are summarized in Table 1.

Table 1. Selected TFPs for the secretion of human interleukin-2

| Number of TFP | Yeast ORF | Number of fused amino acids(total) | Signal sequence | SEQ ID for protein | SEQ ID for DNA |
|---|---|---|---|---|---|
| TFP-9 | YGR106C | 217(265) | Pre(24aa) | 29 | 30 |
| TFP-13 | YIL123W | 127(350) | Pre(19aa) | 31 | 32 |

| | | | | | |
|---|---|---|---|---|---|
| TFP-17 | YNL190W | 68(204) | Pre(20aa) | 33 | 34 |
| TFP-18 | YBR078W | 199(467) | Pre(20aa) | 35 | 36 |
| TFP-19 | YJL178C | 144(271) | Pre(19aa) | 37 | 38 |
| TFP-20 | YMR307W | 187(559) | Pre(22aa) | 39 | 40 |
| TFP-21 | YOR247W | 55(210) | Pre(19aa) | 41 | 42 |
| TFP-25 | YOR085W | 190(350) | Pre(17aa) | 43 | 44 |
| TFP-27 | YKR042W | 89(450) | Pre(17aa) | 45 | 46 |

## EXAMPLE 11

## SECRETION OF HUMAN IL2 USING SELECTED TFPS

**[0125]** To confirm the secretion of human IL2 using selected TFPs, 9 plasmids were constructed using PCR to remove the 5'-UTR of each TFP and *SUC2* of selected plasmids in Example 10 (FIG. 14). Nine forward primers, BamH-YGR-F (SEQ ID NO:47), BamH-SIM-F (SEQ ID NO:48), BamH-YNL-F (SEQ ID NO:49), BamH-ECM-F (SEQ ID NO:50), BamH-ATG-F (SEQ ID NO:51), BamH-GAS-F (SEQ ID NO:52), BamH-YOR-F (SEQ ID NO:53), BamH-OST-F (SEQ ID NO:54), BamH-UTH-F (SEQ ID NO:55) and a common reverse primer IL2-TGA-R (SEQ ID NO:56) were used for PCR from plasmids pYHTS-TFP9, pYHTS-TFP13, pYHTS-TFP17, pYHTS-TFP18, pYHTS-TFP19, pYHTS-TFP20, pYHTS-TFP21, pYHTS-TFP25, and pYHTS-TFP27, respectively. The nine PCR amplified fragments were digested with *Bam*HI and *Sal*I and each fractionated from an agarose gel. Separately, another PCR to amplify the *GAL* promoter was done using a forward primer Sac-GAL-F (SEQ ID NO:57) and a reverse primer GAL-BamH-R (SEQ ID NO:58) from YEGα-HIR525 (Sohn *et al., Process Biochem. 30*:653 (1995)). *Sac*I-*Bam*HI digested *GAL* promoter and the nine *Bam*HI-*Sal*I digested fragments were co-ligated into *Sac*I-*Sal*I digested YEGα-HIR525. The resulting plasmids were named pYGT9-IL2 (FIG. 15A), pYGT13-IL2 (FIG. 15B), pYGT17-IL2 (FIG. 15C), pYGT18-IL2 (FIG. 16A), pYGT19-IL2 (FIG. 16B), pYGT20-IL2 (FIG. 16C), pYGT21-IL2 (FIG. 17A),

pYGT25-IL2 (FIG. 17B), and pYGT27-IL2 (FIG. 17C), respectively. Human IL2 expression vectors, pYGT9-IL2 (*E. coli* DH5α/pYGT9-IL2, FIG. 15A) and pYGT17-IL2 (*E. coli* DH5α/pYGT17-IL2, FIG. 15C) were deposited at an international depository authority, KCTC (Korea Collection for Type Cultures; 52, Oun-dong, Yusong-ku, Taejon, Korea) on July 21, 2005, and assigned accession numbers KCTC 10828BP and KCTC 10829BP, respectively. Nucleotide sequences of all constructed vectors were confirmed to have a correct in-frame fusion between TFP and IL2 and each vector was transformed into *S. cerevisiae* Y2805 (*Mat a ura3 SUC2 pep4::HIS3 GAL1 can1*). Transforming cells were plated on UD media (0.67% yeast nitrogen base without amino acids, 0.77 g/l amino acid mixture, 2% glucose and 2% agar) and incubated at 30°C for 3 days. A single colony of each transformation was inoculated into YPDG broth (1% yeast extract, 2% Bacto-peptone, 1% glucose and 1% galactose) and cultivated at 30°C for 40 hours. Culture supernatant (0.6 ml) was mixed with cold acetone for a final acetone concentration of 40%. After incubation at -20°C for 2 hours, proteins were precipitated by centrifugation for 15 min at 10,000×g. The pellet was freeze-dried and resuspended in 1×SDS-PAGE sample buffer (Bio-Rad, USA) and analyzed in 12% SDS-PAGE. The gel was stained with gel staining reagent (PhastGel® Blue R, Pharmacia Biotech, USA). As shown in FIG. 18, the levels of secreted IL2 were considerably different between the TFPs but all could secrete human IL2 into culture supernatant. A plasmid pYIL-KRT1-4 (WO 2005/068658) containing a TFP1-human IL2 gene was used as a control. TFP9, 13, 21 and 27 were found to be useful for the secretion of human IL2 (FIG. 18).

## EXAMPLE 12

## NOVEL TFP SELECTED FROM TFP LIBRARY FOR THE SECRETION OF HUMAN INTERLEUKIN-32

[0126] As an example for identifying optimal TFPs using a method developed in this invention, a rarely secreting protein, a novel human cytokine, interleukin-32α (hIL32) (Kim *et al., Immunity 22*:131 (2005)) was tested. An insert fragment containing the human IL32α gene and a 500 bp N-terminal part of *SUC2* was amplified using PCR as described in Example 8 (FIG. 11). A PCR was carried out using a forward primer KR-IL32α-F (SEQ ID NO:59) and a reverse primer IL32α-INV-R (SEQ ID NO:60) from pProExHTa-IL32α (DY Yoon, Konkuk University, Korea) as a template. Separately, another PCR for the amplification of an N-terminal part of *SUC2* to be fused to the 3'-end of the IL32α gene was also carried out using a forward primer KR-Inv-F (SEQ ID NO:21) and a reverse primer Inv500-R (SEQ ID NO:22) from YGaINV (FIG. 6). Then, a second PCR was done from the two DNA fragments amplified in the first step using a forward primer LNK40 (SEQ ID NO:23) and a reverse primer Inv500-R (SEQ ID NO:22). The resulting fragment (insert fragment) harbored a 40 nucleotide linker DNA containing a Kex2p recognition sequence (Leu-Asp-Lys-Arg (SEQ ID NO:214)), IL32α, an additional Kex2p recognition sequence, and an N-terminal part of invertase in order. This fragment was co-transformed with the *Swa*I digested TFP library vector constructed in Example 7 into *S. cerevisiae* Y2805Δ*gal1*Δ*suc2* (*Mat a ura3 suc2::Tc190 pep4::HIS3 gal1 can1*). Transformed cells were spread on both UD media (0.67% yeast nitrogen base without amino acids, 0.77 g/l amino acid mixture, 2% glucose and 2% agar) and YPSGA media (1% yeast extract, 2% Bacto-peptone, 2% sucrose, 0.3% galactose, 1 μg/ml antimycin A, and 2% agar) and incubated at 30°C for 5 days. Around $2\times10^4$ transformants were obtained on UD plates but about 250 transformants were obtained on YPSGA. Thirty eight transformants were randomly selected and

cultivated on YPDG broth (1% yeast extract, 2% Bacto-peptone, 1% glucose and 1% galactose) at 30°C for 40 hours. Culture supernatant (0.6 ml) was mixed with cold acetone for a final acetone concentration of 40%. After incubation at -20°C for 2 hours, proteins were precipitated by centrifugation for 15 min at 10,000×g. The pellet was freeze-dried and resuspended in 1×SDS-PAGE sample buffer (Bio-Rad, USA) and analyzed by 12% SDS-PAGE (FIG. 19). Most of the transformants could secrete human IL32α judging from the protein bands appeared around 20 kDa. Among them, 17 transformants showing dark IL32α bands were further analyzed. Each transformant was cultivated on YPD broth and total DNA was isolated and retransformed into *E. coli* DH5α. Transformed *E. coli* was plated on LB media containing ampicillin (1% Bacto-peptone, 0.5% yeast extract, 1% NaCl, with 50 μg/ml ampicillin) and incubated at 37°C overnight. Plasmids were isolated from each *E. coli* transformant using a plasmid extraction kit (Bioneer, Korea). To analyze the sequence of each plasmid, a sequencing primer GAL100-F (SEQ ID NO:12) binding to the *GAL10* promoter was used for all plasmids containing TFPs. Nucleotide sequences were determined by Genotech Co. (Taejon, Korea) using an automated sequencing unit (ABI Prism 377; PE Biosystems, Foster City, CA, USA). The sequences were analyzed by a BLAST search of the *Saccharomyces* Genome Database (www.yeastgenome.org). As a result, nine different TFPs were identified from plasmids isolated from 17 selected yeast strains. The isolated plasmids were named pYHTS-IL32-TFP3, pYHTS-IL32-TFP11, pYHTS-IL32-TFP13, pYHTS-IL32-TFP21, pYHTS-IL32-TFP22, pYHTS-IL32-TFP25, pYHTS-IL32-TFP29 pYHTS-IL32-TFP34, and pYHTS-IL32-TFP38. Among them, TFP3, TFP13, TFP21 and TFP25 were commonly obtained as optimal TFPs for human IL2 (WO 2005/068658) and in Example 10 (Table 1). Five novel TFPs isolated for IL32α are summarized in Table 2.

Table 2. Novel TFPs for the secretion of human interleukin-32α

| Number of TFP | Yeast ORF | Number of fused amino acids(total) | Signal sequence | SEQ ID for protein | SEQ ID for DNA |
|---|---|---|---|---|---|
| TFP-11 | YDR077W | 187(338) | Pre(18aa) | 61 | 62 |
| TFP-22 | YJL159W | 165(310) | PrePro(19+54aa) | 63 | 64 |
| TFP-29 | YEL060C | 48(635) | Pre(19aa) | 65 | 66 |
| TFP-34 | YLR390W-A | 208(238) | Pre(22aa) | 67 | 68 |
| TFP-38 | YMR251W-A | 38(59) | Pre(20aa) | 69 | 70 |

## EXAMPLE 13

## SECRETION OF HUMAN IL32α USING SELECTED TFPS

[0127] To confirm the secretion of human IL32α using selected TFPs, several plasmids were constructed using PCR to remove the 5'-UTR of each TFP and *SUC2* of selected plasmids in Example 12. Six forward primers, BamH-CIS-F (SEQ ID NO:71), BamH-SED-F (SEQ ID NO:72), BamH-SIM-F (SEQ ID NO:73), BamH-YOR247W-F (SEQ ID NO:74), BamH-HSP-F (SEQ ID NO:75), BamH-OST-F (SEQ ID NO:76), and a common reverse primer IL32-TGA-R (SEQ ID NO:77) were used for PCR from plasmids pYHTS-IL32-TFP3, pYHTS-IL32-TFP11, pYHTS-IL32-TFP13, pYHTS-IL32-TFP21, pYHTS-IL32-TFP22, and pYHTS-IL32-TFP25, respectively. The six PCR amplified fragments were digested with *Bam*HI and *Sal*I and each fractionated from an agarose gel. Separately, another PCR to amplify the *GAL* promoter was done using a forward primer Sac-GAL-F (SEQ ID NO:57) and a reverse primer GAL-BamH-R (SEQ ID NO:58) from YEGα-HIR525 (Sohn *et al.*, *Process Biochem. 30*:653 (1995)). *Sac*I-*Bam*HI digested *GAL* promoter and the six *Bam*HI-*Sal*I digested fragments were co-ligated into *Sac*I-*Sal*I digested YEGα-HIR525. The resulting plasmids were named pYGT3-IL32α, pYGT11-IL32α, pYGT13-IL32α, pYGT21-IL32α, pYGT22-IL32α, and pYGT25-IL32α, respectively. Nucleotide sequences of all constructed vectors were confirmed to have a correct in-frame fusion between TFP and IL32α and

each vector was transformed into *S. cerevisiae* Y2805 (*Mat a ura3 SUC2 pep4::HIS3 GAL1 can1*). Transformed cells were plated on UD media (0.67% yeast nitrogen base without amino acids, 0.77 g/l amino acid mixture, 2% glucose and 2% agar) and incubated at 30°C for 3 days. A single colony of each transformation was inoculated into YPDG broth (1% yeast extract, 2% Bacto-peptone, 1% glucose and 1% galactose) and cultivated at 30°C for 40 hours. Culture supernatant (0.6 ml) was mixed with cold acetone for a final acetone concentration of 40%. After incubation at -20°C for 2 hours, proteins were precipitated by centrifugation for 15 min at 10,000×g. The pellet was freeze-dried and resuspended in 1×SDS-PAGE sample buffer (Bio-Rad, USA) and analyzed on 12% SDS-PAGE. The gel was stained with gel staining reagent (PhastGel® Blue R, Pharmacia Biotech, USA). Secreted IL32α was further analyzed by Western blotting using a monoclonal antibody of hIL32α. Proteins were transferred to PVDF membranes (Millipore, USA) in CAPS buffer (2.2 g per liter CAPS, MeOH 10%, pH 11 adjusted with NaOH) using a Mighty small tank transfer (Hoefer, USA) at 300 mA for 90 min. Proteins were then detected with human IL32 antibody (DY Yoon, Konkuk University, Korea). Membranes were blocked overnight at 4°C in PBS (137 mM NaCl, 2.7 mM KCl, 10 mM $Na_2HPO_4$, 2 mM $KH_2PO_4$, pH 7.4 adjusted with HCl) containing 5% skim milk. Membranes were washed 3 times with PBS containing 0.05% Tween-20 and then incubated with primary antibodies diluted in PBS containing 3% skim milk at room temperature for 1 hour. Membranes were then washed 3 times and incubated with the anti-mouse secondary antibody (Sigma Chemical Co., USA) diluted in PBS containing 3% skim milk at room temperature for 1 hour. Membranes were washed as above and developed with Sigma Fast NBT/BCIP (Sigma Chemical Co., USA). As shown in FIG. 20, all selected TFPs could secrete human IL32α into the culture supernatant. Among them, TFP3, 13, 21 and 22 were found to be optimal for the secretion of human IL32α.

## EXAMPLE 14

## FED-BATCH FERMENTATION FOR THE PRODUCTION OF HUMAN IL32α

[0128] A recombinant yeast strain transformed with pYGT3-IL32α was cultured in a 5-L jar fermentor by fed-batch culture for the evaluation of the secretory productivity of human IL32α. A 200 ml seed culture was cultured in a 1 liter flask using a minimal medium (0.67% yeast nitrogen base without amino acids, 0.5% casamino acids and 2% glucose). When the culture using a fermentation culture medium (4% yeast extract, 1% peptone, 2% glucose) as an initial fermentation medium reached an OD600 of about 15, a fed-batch medium (15% yeast extract, 30% glucose) was supplied with different feeding rates according to cell growth rates. After the culture reached an OD600 of about 130, galactose (30% galactose) was additionally supplied with different feeding rates according to cell growth rates. After a culture period of about 72 hrs, the culture reached an OD600 of about 220 (FIG. 21A). 15 μl of the medium was collected at the given time points and assessed for secreted proteins by SDS-PAGE (FIG. 21B). Over 300 mg/L of hIL32α was found to be secreted into the culture medium as determined by the direct measurement of proteins with BCA protein assay reagent (Pierce, USA) and with a densitometer (GS700, Bio-Rad, USA).

## EXAMPLE 15

## SEQUENCE-BASED SELECTION OF TFPS USING BLAST SEARCH FROM YEAST GENOMIC DATABASE

[0129] For the sequence-based selection of TFPs from the yeast genome, amino acid sequences of pre-secretion signals of 18 selected TFPs (4 from WO 2005/068658, 9 from example 10 and 5 from example 12) were used as a query sequence for a BLAST search of the *Saccharomyces* Genome Database

(www.yeastgenome.org). Using a low expect threshold (100 or 1000) in the BLASTP search, several hundred ORFs having over 70% homology were identified. Of those, the ORFs with sequence homology near the N-terminus were selected, and further subjected to SignalP (www.cbs.dtu.dk/services/SignalP-2.0/) analysis for the selection of ORFs with secretion signal. As a result, 18 ORFs were randomly selected as TFP candidates. Eighteen selected ORFs identified by the search were YGR279C (SCW4, cell wall protein), YLR037C (DAN2, cell wall mannoprotein), YLR110C (CCW12, cell wall protein), YOR383C (FIT3, cell wall mannoprotein), YIL011W (TIR3, cell wall mannoprotein), YHR214W (putative membrane protein), YNL160W (YGP1, cell wall-related secretory glycoprotein), YGR296C-A (dubious open reading frame), YOL154W (ZPS1, putative GPI-anchored protein), YPL187W (MFα, mating pheromone alpha-factor), YHR214W (putative membrane protein), YKR013W (PRY2, protein of unknown function), YHR139C (SPS100, protein required for spore wall maturation), YIL169C (putative protein of unknown function), YOL155C (uncharacterized ORF), YMR325W (PAU19, hypothetical protein), YDR134W (hypothetical protein) and YLR300W (EXG1, major exo-1,3-beta-glucanase of the cell wall). Each ORF was amplified from the genomic DNA of *S. cerevisiae* Y2805 (*Mat a ura3 SUC2 pep4::HIS3 GAL1 can1*) using PCR primer pairs YGR279C-F(SEQ ID NO:92) and YGR279C-R (SEQ ID NO:93) for YGR279C, YLR037C-F (SEQ ID NO:94) and YLR037C-R (SEQ ID NO:95) for YLR037C, YLR110C-F (SEQ ID NO:96) and YLR110C-R (SEQ ID NO:97) for YLR110C, YOR383C-F (SEQ ID NO:98) and YOR383C-R (SEQ ID NO:99) for YOR383C, YIL011W-F (SEQ ID NO:100) and YIL011W-R (SEQ ID NO:101) for YIL011W, YHR214W-F (SEQ ID NO:102) and YHR214W-R (SEQ ID NO:103) for YHR214W, YNL160W-F (SEQ ID NO:104) and YNL160W-R (SEQ ID NO:105) for YNL160W, YGR296C-A-F (SEQ ID NO:106) and YGR296C-A-R (SEQ ID NO:107) for YGR296C-A, YOL154W-F (SEQ ID NO:108) and YOL154W-R (SEQ ID NO:109) for YOL154W, YPL187W-F (SEQ ID NO:110) and YPL187W-R

(SEQ ID NO:111) for YPL187W, YHR214W-F (SEQ ID NO:112) and YHR214W-R (SEQ ID NO:113) for YHR214W, YKR013W-F (SEQ ID NO:114) and YKR013W-R (SEQ ID NO:115) for YKR013W, YHR139C-F (SEQ ID NO:116) and YHR139C-R (SEQ ID NO:117) for YHR139C, YIL169C-F (SEQ ID NO:118) and YIL169C-R (SEQ ID NO:119) for YIL169C, YOL155C-F (SEQ ID NO:120) and YOL155C-R (SEQ ID NO:121) for YOL155C, YMR325W-F (SEQ ID NO:122) and YMR325W-R (SEQ ID NO:123) for YMR325W, YDR134W-F (SEQ ID NO:124) and YDR134W-R (SEQ ID NO:125) for YDR134W and YLR300W-F (SEQ ID NO:126) and YLR300W-R (SEQ ID NO:127) for YLR300W, respectively. PCR was performed with *Pfu* polymerase (Stratagene, USA) and PCR conditions included one cycle of 94°C for 3 min, and 25 cycles of 94°C for 30 sec, 55°C for 30 sec and 72°C for 2 min, followed by one final cycle of 72°C for 7 min. Each amplified PCR fragment was confirmed by nucleotide sequencing by Genotech Co. (Taejon, Korea) using an automated sequencing unit (ABI Prism 377; PE Biosystems, Foster City, CA, USA).

[0130] For the screening of TFPs from the selected 18 ORFs, unidirectional deletion of the mixture of 18 PCR fragments was carried out and used for the construction of a TFP library in YGadV45 (FIG. 24). Single stranded template was obtained by unidirectional PCR using a primer SfiA-F (SEQ ID NO:128) from the template consisting of the 18 ORFs. PCR was performed with ExTaq (Takara Korea, Korea) and PCR conditions included one cycle of 94°C for 3 min, and 30 cycles of 94°C for 30 sec, 55°C for 30 sec and 72°C for 2 min, followed by one final cycle of 72°C for 7 min. PCR product containing the single stranded DNA was purified using a PCR purification kit (Bioneer, Korea). Then, the regeneration of double stranded DNA was carried out using *E. coli* DNA polymerase I (NEB, England) and a random hexameric primer, ASA24N6 (SEQ ID NO:16). A reaction mixture containing 20 μl of template DNA, 1 μl of ASA24N6 primer, 3 μl of 10x *E. coli* DNA polI buffer, 5 μl of 2.5 mM dNTP, and 1 μl of *E. coli* DNA polI was incubated at 37°C for 1 hour. The DNA was column purified using a PCR purification kit (Bioneer,

Korea) and PCR amplified using primers SfiA-F (SEQ ID NO:128) and ASA24 (SEQ ID NO:17). The amplified DNA was column purified again, digested with *Sfi*I and fractionated by agarose gel electrophoresis. 0.5-1.0 kb of *Sfi*I digested DNA was subcloned into the *Sfi*I treated YGadV45 containing a defective *SUC2* (d*SUC2*). The ligated DNA was transformed into *E. coli* DH5α. Transformed *E. coli* was plated on LB media with ampicillin (1% Bacto-peptone, 0.5% yeast extract, 1% NaCl with 50 μg/ml ampicillin) and incubated at 37°C overnight. About $1\times10^4$ *E. coli* colonies were pooled with sterile distilled water and the total plasmids containing the unidirectional-deleted DNA fragment library of 18 ORFs in YGadV45, were isolated by using a plasmid isolation kit (Bioneer, Korea).

[0131] To screen proper TFPs from the unidirectional-deleted DNA fragment library of 18 ORFs, a gene encoding human interleukin-2 (hIL2) was inserted between the library and *dSUC2*. An insert fragment containing the hIL2 gene and a 500 bp N-terminal part of *SUC2* was amplified using PCR as described in Example 8 (FIG. 11). This fragment was co-transformed with *Swa*I digested vector containing the unidirectional-deleted DNA fragment library of 18 ORFs into *S. cerevisiae* Y2805Δ*gal1*Δ*suc2* (*Mat a ura3 suc2::Tc190 pep4::HIS3 gal1 can1*). Transformed cells were spread on both UD media (0.67% yeast nitrogen base without amino acids, 0.77 g/l amino acid mixture, 2% glucose and 2% agar) and YPSGA media (1% yeast extract, 2% Bacto-peptone, 2% sucrose, 0.3% galactose, 1 μg/ml antimycin A, and 2% agar) and incubated at 30°C for 5 days. Around $2\times10^4$ transformants were obtained on UD plates but about several hundred transformants were obtained in YPSGA. A random selection of 29 transformants growing on YPSGA was cultivated on YPDG broth (1% yeast extract, 2% Bacto-peptone, 1% glucose and 1% galactose) and cultivated at 30°C for 40 hours. Culture supernatant (0.6 ml) was mixed with cold acetone for a final acetone concentration of 40%. After incubation at -20°C for 2 hours, proteins were precipitated by centrifugation for 15 min at 10,000×g. The pellet was freeze-dried and resuspended in 1×SDS-PAGE sample buffer (Bio-Rad, USA) and analyzed by 12% SDS-

PAGE (FIG. 25). Several transformants were found to secrete human IL-2 into the culture supernatant. Total DNA was isolated from each cell secreting human IL-2 and retransformed into *E. coli* DH5α. Plasmids were isolated from each *E. coli* transformant using a plasmid extraction kit (Bioneer, Korea). To analyze the sequence of each TFP, a sequencing primer GAL100-F (SEQ ID NO:12) binding to the *GAL10* promoter was used for all plasmids containing TFPs. Nucleotide sequences were determined by Genotech Co. (Taejon, Korea) using an automated sequencing unit (ABI Prism 377; PE Biosystems, Foster City, CA, USA). The sequences were analyzed by a BLAST search of the *Saccharomyces* Genome Database (www.yeastgenome.org). As a result, six novel TFPs were identified from plasmids isolated from the 12 transformants secreting human IL-2. The isolated plasmids were named pYIL-TFP39, pYIL-TFP41, pYIL-TFP43, pYIL-TFP44, pYIL-TFP52, and pYIL-TFP54, respectively. The six novel TFPs are summarized in Table 3.

Table 3. TFPs from sequence-based selected ORFs for the secretion of human IL-2

| Number of TFP | Yeast ORF | Number of fused amino acids(total) | Signal sequence | SEQ ID for protein | SEQ ID for DNA |
|---|---|---|---|---|---|
| TFP-39 | YGR279C | 57(386) | Pre(19aa) | 129 | 130 |
| TFP-43 | YLR110C | 129(133) | Pre(18aa) | 131 | 132 |
| TFP-44 | YOR383C | 71(204) | Pre(18aa) | 133 | 134 |
| TFP-48 | YGR279C | 119(386) | Pre(19aa) | 135 | 136 |
| TFP-52 | YNL160W | 129(354) | Pre(19aa) | 137 | 138 |
| TFP-54 | YLR037C | 124(124) | Pre(20aa) | 139 | 140 |

## EXAMPLE 16

## DIVERSIFICATION OF CORE-TFPS BY UNIDIRECTIONAL DELETION

[0132] To diversify the usefulness of 14 TFPs (core-TFPs) selected by using IL-2 and IL-32α in Examples 10 and 11, and 3 TFPs previously identified in WO 2005/068658, seventeen genomic ORFs, YAR066W for TFP-1, YFR026C for TFP-2, YJL158C for TFP3, YGR106C for TFP-9, YDR077W for TFP-11, YIL123W for TFP13, YNL190W for TFP-17, YBR078W for TFP18, YJL178C for TFP-19, YMR307W for TFP-20, YOR247W for TFP-21, YJL159W for TFP-22, YOR085W for TFP-25, YKR042W for TFP-27, YEL060C for TFP29, YLR390W-A for TFP-34, and YMR251W-A for TFP-38, were PCR amplified and unidirectionally deleted as described in Example 15. Each ORF was amplified from the genomic DNA of *S. cerevisiae* Y2805 (*Mat a ura3 SUC2 pep4::HIS3 GAL1 can1*) using PCR primer pairs YAR066W-F (SEQ ID NO:141) and YAR066W-R (SEQ ID NO:142) for YAR066W, YFR026C-F (SEQ ID NO:143) and YFR026C-R (SEQ ID NO:144) for YFR026C, YJL158C-F (SEQ ID NO:145) and YJL158C-R (SEQ ID NO:146) for YJL158C, YGR106C-F (SEQ ID NO:147) and YGR106C-R (SEQ ID NO:148) for YGR106C, YDR077W-F (SEQ ID NO:149) and YDR077W-R (SEQ ID NO:150) for YDR077W, YIL123W-F (SEQ ID NO:151) and YIL123W-R (SEQ ID NO:152) for YIL123W, YNL190W-F (SEQ ID NO:153) and YNL190W-R (SEQ ID NO:154) for YNL190W, YBR078W-F (SEQ ID NO:155) and YBR078W-R (SEQ ID NO:156) for YBR078W, YJL178C-F (SEQ ID NO:157) and YJL178C-R (SEQ ID NO:158) for YJL178C, YMR307W-F (SEQ ID NO:159) and YMR307W-R (SEQ ID NO:160) for YMR307W, YOR247W-F (SEQ ID NO:161) and YOR247W-R (SEQ ID NO:162) for YOR247W, YJL159W-F (SEQ ID NO:163) and YJL159W-R (SEQ ID NO:164) for YJL159W, YOR085W-F (SEQ ID NO:165) and YOR085W-R (SEQ ID NO:166) for YOR085W, YKR042W-F (SEQ ID NO:167) and YKR042W-R (SEQ ID NO:168) for

YKR042W, YEL060C-F (SEQ ID NO:169) and YEL060C-R (SEQ ID NO:170) for YEL060C, YLR390W-A-F (SEQ ID NO:171) and YLR390W-A-R (SEQ ID NO:172) for YLR390W-A, YMR251W-A-F (SEQ ID NO:173) and YMR251W-A-R (SEQ ID NO:174) for YMR251W-A, respectively. PCR was performed with *Pfu* polymerase (Stratagene, USA) and PCR conditions included one cycle of 94°C for 3 min, and 25 cycles of 94°C for 30 sec, 55°C for 30 sec and 72°C for 2 min, followed by one final cycle of 72°C for 7 min. Each amplified PCR fragment was confirmed by nucleotide sequencing by Genotech Co. (Taejon, Korea) using an automated sequencing unit (ABI Prism 377; PE Biosystems, Foster City, CA, USA).

[0133] For the screening of diversified TFPs from the 17 ORFs from which 17 core-TFPs were obtained, unidirectional deletion of the mixture of 17 PCR fragments was carried out and used for the construction of a TFP library in YGadV45 (FIG. 24). Single stranded template was obtained by unidirectional PCR using a primer SfiA-F (SEQ ID NO:128) from the template consisting of 17 ORFs. PCR was performed with ExTaq (Takara Korea, Korea) and PCR conditions included one cycle of 94°C for 3 min, and 30 cycles of 94°C for 30 sec, 55°C for 30 sec and 72°C for 2 min, followed by one final cycle of 72°C for 7 min. PCR product containing the single stranded DNA was purified using a PCR purification kit (Bioneer, Korea). Then, the regeneration of double stranded DNA was carried out using *E. coli* DNA polymerase I (NEB, England) and a random hexameric primer, ASA24N6 (SEQ ID NO:16). A reaction mixture containing 20 μl of template DNA, 1 μl of ASA24N6 primer, 3 μl of 10x *E. coli* DNA polI buffer, 5 μl of 2.5 mM dNTP, and 1 μl of *E. coli* DNA polI was incubated at 37°C for 1 hour. The DNA was column purified using a PCR purification kit (Bioneer, Korea) and PCR amplified using primers SfiA-F (SEQ ID NO:128) and ASA24 (SEQ ID NO:17). The amplified DNA was column purified again, digested with *Sfi*I and fractionated by agarose gel electrophoresis. 0.5-1.0 kb of *Sfi*I digested DNA was subcloned into *Sfi*I treated YGadV45 containing a defective *SUC2* (d*SUC2*). The ligated DNA was transformed into *E. coli* DH5α. Transformed *E. coli*

was plated on LB media with ampicillin (1% Bacto-peptone, 0.5% yeast extract, 1% NaCl with 50 μg/ml ampicillin) and incubated at 37°C overnight. About $1\times10^4$ *E. coli* colonies were pooled with sterile distilled water and the total plasmids containing the unidirectional-deleted DNA fragment library of 17 ORFs in YGadV45 were isolated by using a plasmid isolation kit (Bioneer, Korea). Two unidirectional-deleted library DNAs from 17 ORFs for core-TFPs and 18 ORFs prepared in Example 15 were combined for further application.

[0134] To screen proper TFPs from the unidirectional-deleted DNA fragment library from 35 ORFs, a gene encoding human interleukin-2 (hIL2) was inserted between the library and *dSUC2*. An insert fragment containing the human IL2 gene and a 500 bp N-terminal part of *SUC2* was amplified using PCR as described in Example 8 (FIG. 11). This fragment was co-transformed with *Swa*I digested vector containing the unidirectional-deleted DNA fragment library of 35 ORFs, into *S. cerevisiae* Y2805Δ*gal1*Δ*suc2* (*Mat a ura3 suc2::Tc190 pep4::HIS3 gal1 can1*). Transformed cells were spread on both UD media (0.67% yeast nitrogen base without amino acids, 0.77 g/l amino acid mixture, 2% glucose and 2% agar) and YPSGA media (1% yeast extract, 2% Bacto-peptone, 2% sucrose, 0.3% galactose, 1 μg/ml antimycin A, and 2% agar) and incubated at 30°C for 5 days. Around $2\times10^4$ transformants were obtained on UD plates but about several hundred transformants were obtained in YPSGA. A random selection of 24 transformants growing on YPSGA was cultivated on YPDG broth (1% yeast extract, 2% Bacto-peptone, 1% glucose and 1% galactose) and cultivated at 30°C for 40 hours. Culture supernatant (0.6 ml) was mixed with cold acetone for a final acetone concentration of 40%. After incubation at -20°C for 2 hours, proteins were precipitated by centrifugation for 15 min at 10,000×g. The pellet was freeze-dried and resuspended in 1×SDS-PAGE sample buffer (Bio-Rad, USA) and analyzed by 12% SDS-PAGE (FIG. 26). Most of the transformants could secrete human IL-2 into the culture supernatant but with different levels between them. Total DNA was isolated from each transformant secreting human IL-2 and

retransformed into *E. coli* DH5α. Plasmids were isolated from *E. coli* using a plasmid extraction kit (Bioneer, Korea). To analyze the sequence of each TFP, a sequencing primer GAL100-F (SEQ ID NO:12) binding to the *GAL10* promoter were used for all plasmids containing TFPs. Nucleotide sequences were determined by Genotech Co. (Taejon, Korea) using an automated sequencing unit (ABI Prism 377; PE Biosystems, Foster City, CA, USA). The sequences were analyzed by a BLAST search of the *Saccharomyces* Genome Database (www.yeastgenome.org). As a result, six novel TFPs were identified from plasmids isolated from the 18 transformants secreting human IL-2. The isolated plasmids were named pYIL-TFP40, pYIL-TFP50, pYIL-TFP51, pYIL-TFP57, pYIL-TFP58, and pYIL-TFP59, respectively. The six novel TFPs are summarized in Table 4.

Table 4. TFPs from sequence-based selected ORFs for the secretion of human IL-2

| Number of TFP | Yeast ORF | Number of fused amino acids(total) | Signal sequence | SEQ ID for protein | SEQ ID for DNA |
|---|---|---|---|---|---|
| TFP-40 | YGR279C | 99(386) | Pre(19aa) | 175 | 176 |
| TFP-50 | YOR247W | 85(210) | Pre(19aa) | 177 | 178 |
| TFP-51 | YOR247W | 116(210) | Pre(19aa) | 179 | 180 |
| TFP-57 | YOL155C | 114(967) | Pre(23aa) | 181 | 182 |
| TFP-58 | YAR066W | 199(203) | Pre(23aa) | 183 | 184 |
| TFP-59 | YOR085W | 55(350) | Pre(17aa) | 185 | 186 |

## EXAMPLE 17

## ARTIFICIAL TFPS USING SWAPPING OF PRE AND PRO SIGNAL SEQUENCE BETWEEN CORE-TFPS

[0135] To date, a yeast secretion signal from mating factor alpha (MFα) has been the most widely used for the secretion of various recombinant proteins in yeast (Romanos *et al., Yeast 8*:423 (1992)). The secretion signal comprises 19 amino acids of pre-signal and 66 amino acids of pro-signal. The exact

function of pro-signal is uncertain but it has been known to be essential for the correct folding and secretion of some proteins. The fact was also investigated in the secretion of some recombinant proteins in yeast (Chaudhuri *et al., Eur. J. Biochem. 206*:793 (1992)). In this invention, two secretion signals, TFP-3 and TFP-22, were identified as pre-pro type. For the expansion of the usefulness of TFPs selected in this invention, artificial TFPs were designed to have a different origin of the pre and pro signals. Four artificial TFPs were constructed using the pre-signal of TFP-1, 2, 3 and 4 and a common pro-signal of mating factor alpha and the resulting TFPs were named as TFP-5, 6, 7, and 8. For the fusion between 4 different pre-signals and a common pro signal, overlap extension PCR was used.

[0136] A first step PCR was carried out for the amplification of four different pre-signals of 4 TFPs using primer pairs T1-F (SEQ ID NO:187) and T1-R (SEQ ID NO:188), T2-F (SEQ ID NO:189) and T2-R (SEQ ID NO:190), T3-F (SEQ ID NO:191) and T3-R (SEQ ID NO:192), T4-F (SEQ ID NO:193) and T4-R (SEQ ID NO:194) from plasmids pYIL-KRTFP1, 2, 3, and 4 (WO 2005/068658), respectively. Separately, another PCR for the amplification of about 190 bp of mating factor alpha pro-signal was also carried out using primers MF-Pro-F (SEQ ID NO:195) and MF-R (SEQ ID NO:196) from plasmid YEGα-HIR525. Then second PCRs for the 4 different pre-pro signals were done from 4 sets of two DNA fragments, 4 pre-signals and a MFα pro-signal amplified in the first step using 4 different forward primers, T1-F (SEQ ID NO:187), T2-F (SEQ ID NO:189), T3-F (SEQ ID NO:191) and T4-F (SEQ ID NO:193) and a common reverse primer, MF-R (SEQ ID NO:196), respectively. To compare the efficiency of each artificial pre-pro signal sequence with that of mating factor alpha, pre-pro signal of mating factor alpha was also PCR amplified using primers MF-Pre-F (SEQ ID NO:197) and MF-R (SEQ ID NO:196) from YEGα-HIR525.

[0137] A target protein, human insulin-like growth factor (hIGF) was selected to test the five pre-pro signal sequences. It has been reported that the pro signal of mating factor alpha was necessary for the secretion of human insulin-

like growth factor in yeast (Chaudhuri *et al., Eur. J. Biochem. 206*:793 (1992)). Human IGF gene was first PCR amplified using primers KR-IGF-F (SEQ ID NO:198) and IGF-R (SEQ ID NO:199) from a human cDNA library (ES Choi, Korea Research Institute of Bioscience and Biotechnology, Korea) and then a second PCR was done using LNK40 (SEQ ID NO:23) and IGF-R (SEQ ID NO:199). The DNA fragment containing IGF was fused to the previously amplified 5 PCR fragments containing pre-pro signals using 5 forward primers, T1-F (SEQ ID NO:187), T2-F (SEQ ID NO:189), T3-F (SEQ ID NO:191), T4-F (SEQ ID NO:193), MF-Pre-F (SEQ ID NO:197) and a common reverse primer IGF-R (SEQ ID NO:199). All fused PCR products were digested with *Sfi*I and *Sal*I and then subcloned into the *Sfi*I-*Sal*I digested vector YGalINV (FIG. 6). The resulting plasmids were named pYGa-T1α-IGF, pYGa-T2α-IGF pYGa-T3α-IGF pYGa-T4α-IGF and pYGa-MFα-IGF, respectively. Five plasmids were transformed into *S. cerevisiae* Y2805 (*Mat a ura3 pep4::HIS3 gal1 can1*). Transformed cells were spread on UD media (0.67% yeast nitrogen base without amino acids, 0.77 g/l amino acid mixture, 2% glucose and 2% agar). A single colony of each transformation was isolated and cultivated in YPDG (1% yeast extract, 2% Bacto-peptone, 1% glucose and 1% galactose) at 30°C for 40 hours. Culture supernatant (0.6 ml) was mixed with cold acetone for a final acetone concentration of 40%. After incubation at -20°C for 2 hours, proteins were precipitated by centrifugation for 15 min at 10,000×g. The pellet was freeze-dried and resuspended in 1×SDS-PAGE sample buffer (Bio-Rad, USA) and analyzed by 12% SDS-PAGE. Secreted IGF was further analyzed by Western blotting using an antibody for hIGF (FIG 27). All tested pre-pro secretion signals could secrete human IGF into the culture supernatant but with different efficiencies. Among 5 pre-pro signals, T3α (pre-signal from TFP-3 and pro-signal from MFα) and T4α (pre-signal from TFP-4 and pro-signal from MFα) were found to be effective for the secretion of human IGF. The four artificial TFPs and a novel TFP are summarized in Table 5.

Table 5. Novel TFPs for the secretion of human IGF

| Number of TFP | Yeast ORF | Number of fused amino acids | Signal sequence | SEQ ID for protein | SEQ ID for DNA |
|---|---|---|---|---|---|
| TFP-5 | YAR066W/YPL187W | 88 | PrePro(23+65aa) | 200 | 201 |
| TFP-6 | YFR026C/YPL187W | 84 | PrePro(19+65aa) | 202 | 203 |
| TFP-7 | YJL158C/YPL187W | 86 | PrePro(21+65aa) | 204 | 205 |
| TFP-8 | HpPRB1/YPL187W | 83 | PrePro(18+65aa) | 206 | 207 |
| TFP-32 | YPL187W | 84 | PrePro(19+65aa) | 208 | 209 |

## EXAMPLE 18

## CONSTRUCTION OF THE SELECTED TFP VECTORS APPLICABLE TO MANY TARGET GENES THROUGH *IN VIVO* RECOMBINATION

**[0138]** Thirty five TFPs (core-TFPs) selected in this invention (4 TFPs from WO 2005/068658, 14 TFPs selected using two reporter proteins, human IL2 and IL32α in Example 10 and 11, 6 TFPs from ORFs selected by BLAST search in Example 15, 6 TFPs from unidirectional deletion of ORFs encoding the pre-selected TFPs in Example 16, 5 TFPs from artificial design of TFPs in Example 17) might be also useful for the secretion of other proteins. To apply such vectors to large numbers of target genes, the core-TFP vectors were reconstructed for *in vivo* recombination with target genes. For the construction of plasmid YGaSW, a PCR for the amplification of 170 bp fragment containing an *Eco*RI, 2 *Sfi*I, *Not*I, a linker DNA containing a Kex2p recognition site, *Swa*I and *Sal*I site was carried out using primers GAL100-F (SEQ ID NO:12) and H77-1-R (SEQ ID NO:78) from YGadV45 (FIG. 10). An *Eco*RI-*Sal*I digested PCR fragment was subcloned into *Eco*RI-*Sal*I digested YGadV45 and the resulting plasmid was named YGaSW. The plasmid harbors restriction sites for *Eco*RI, *Sfi*I, *Not*I, *Sfi*I, a 40 bp linker and restriction sites *Swa*I and *Sal*I between the *GAL10* promoter and the *GAL7* terminator. Thirty five core-TFPs were obtained by the *Sfi*I digestion of plasmids containing each TFP. Each core-TFP was gel purified and subcloned

into *Sfi*I digested YGaSW and the resulting 35 plasmids were named YGaSW-TFP1, YGaSW-TFP2, YGaSW-TFP3, YGaSW-TFP4, YGaSW-TFP5, YGaSW-TFP6, YGaSW-TFP7, YGaSW-TFP8, YGaSW-TFP9, YGaSW-TFP11, YGaSW-TFP13, YGaSW-TFP17, YGaSW-TFP18, YGaSW-TFP19, YGaSW-TFP20, YGaSW-TFP21, YGaSW-TFP22, YGaSW-TFP25, YGaSW-TFP27, YGaSW-TFP29, YGaSW-TFP32 YGaSW-TFP34, YGaSW-TFP38, YGaSW-TFP39, YGaSW-TFP40, YGaSW-TFP43, YGaSW-TFP44, YGaSW-TFP48, YGaSW-TFP50, YGaSW-TFP51, YGaSW-TFP52, YGaSW-TFP54, YGaSW-TFP57, YGaSW-TFP58, and YGaSW-TFP59, respectively.

## EXAMPLE 19

## EVALUATION OF SELECTED CORE-TFPS FOR THE SECRETION OF HUMAN GROWTH HORMONE

[0139] Core-TFPs selected in this invention were tested for the secretion of human growth hormone (hGH). The human GH gene was PCR amplified from a human cDNA library (ES Choi, Korea Research Institute of Bioscience and Biotechnology, Korea) using primers hGH-F (SEQ ID NO:79) and hGH-R (SEQ ID NO:80) and subcloned into pST-Blue1 (Novagen, USA). The resulting plasmid was named pST-hGH. A second PCR was carried out using primers KR-hGH-F (SEQ ID NO:81) and hGH-Sal-R (SEQ ID NO:82) from pST-hGH. The PCR product containing the hGH gene was used for a third PCR using primers LNK40 (SEQ ID NO:23) and GT70-R (SEQ ID NO:83) to add homologous sequences with YGaSW-TFP vectors constructed in Example 18. The amplified PCR fragment was mixed 2:1 with *Swa*I digested YGaSW-TFP vectors and transformed into *S. cerevisiae* Y2805 (*Mat a ura3 SUC2 pep4::HIS3 GAL1 can1*) through *in vivo* recombination. Transformed cells were plated on UD media (0.67% yeast nitrogen base without amino acids, 0.77 g/l amino acid mixture, 2% glucose and 2% agar) and incubated at 30°C for 3 days. A single colony of each transformation was inoculated into YPDG

broth (1% yeast extract, 2% Bacto-peptone, 1% glucose and 1% galactose) and cultivated at 30°C for 40 hours. Culture supernatant (0.6 ml) was mixed with cold acetone for a final acetone concentration of 40%. After incubation at -20°C for 2 hours, proteins were precipitated by centrifugation for 15 min at 10,000×g. The pellet was freeze-dried and resuspended in 1×SDS-PAGE sample buffer (Bio-Rad, USA) and analyzed by 12% SDS-PAGE. As shown in FIG. 22, most TFPs could secrete human growth hormone into the culture supernatant. Among them, a strain with pYGT21-hGH was tested for the secretion level during fed-batch fermentation. Ten microliters of culture supernatant sampled at the indicated time points were analyzed by SDS-PAGE (FIG. 23). Around 500 mg/liter of human growth hormone was secreted into the culture supernatant.

## EXAMPLE 20

## EVALUATION OF SELECTED CORE-TFPS FOR THE SECRETION OF HUMAN CASPASE-1 SUBUNIT P10

**[0140]** Core-TFPs selected in this invention were tested for the secretion of human caspase-1 subunit p10(hP10). The human p10 gene was PCR amplified from a human cDNA library (ES Choi, Korea Research Institute of Bioscience and Biotechnology, Korea) using primers KR-hP10-F (SEQ ID NO:210) and hP10-Sal-R (SEQ ID NO:211). The PCR product containing the hP10 gene was used for a second PCR using primers LNK40 (SEQ ID NO:23) and GT70-R (SEQ ID NO:83) to add homologous sequences with YGaSW-TFP vectors constructed in Example 18. The amplified PCR fragment was mixed 2:1 with *Swa*I digested YGaSW-TFP vectors and transformed into *S. cerevisiae* Y2805 (*Mat a ura3 SUC2 pep4::HIS3 GAL1 can1*) through *in vivo* recombination. Transformed cells were plated on UD media (0.67% yeast nitrogen base without amino acids, 0.77 g/l amino acid mixture, 2% glucose and 2% agar) and incubated at 30°C for 3 days. A single colony of each

transformation was inoculated into YPDG broth (1% yeast extract, 2% Bacto-peptone, 1% glucose and 1% galactose) and cultivated at 30°C for 40 hours. Culture supernatant (0.6 ml) was mixed with cold acetone for a final acetone concentration of 40%. After incubation at -20°C for 2 hours, proteins were precipitated by centrifugation for 15 min at 10,000×g. The pellet was freeze-dried and resuspended in 1×SDS-PAGE sample buffer (Bio-Rad, USA) and analyzed by 12% SDS-PAGE. As shown in FIG. 28, only 4 artificial TFPs containing pre-pro signals could secrete hP10 protein into the culture supernatant. As found in the case of hIGF, pro-signal was necessary for the proper secretion of human caspase-1 subunit P10 in yeast.

## EXAMPLE 21

## EVALUATION OF SELECTED CORE-TFPS FOR THE SECRETION OF HUMAN INTERLEUKIN-32γ

[0141] Core-TFPs selected in this invention were tested for the secretion of human interleukin-32γ (hIL32γ). A gene coding for human interleukin 32 splicing variant gamma was PCR amplified from pGMT-IL32γ (DY Yoon, Konkuk University, Korea) using primers KR-hIL32g-F (SEQ ID NO:212) and hIL32g-Sal-R (SEQ ID NO:213). The PCR product containing the hIL32γ gene was used for a second PCR using primers LNK40 (SEQ ID NO:23) and GT70-R (SEQ ID NO:83) to add homologous sequences with YGaSW-TFP vectors constructed in Example 18. The amplified PCR fragment was mixed 2:1 with *Swa*I digested YGaSW-TFP vectors and transformed into *S. cerevisiae* Y2805 (*Mat a ura3 SUC2 pep4::HIS3 GAL1 can1*) through *in vivo* recombination. Transformed cells were plated on UD media (0.67% yeast nitrogen base without amino acids, 0.77 g/l amino acid mixture, 2% glucose and 2% agar) and incubated at 30°C for 3 days. A single colony of each transformation was inoculated into YPDG broth (1% yeast extract, 2% Bacto-peptone, 1% glucose and 1% galactose) and cultivated at

30°C for 40 hours. Culture supernatant (0.6 ml) was mixed with cold acetone for a final acetone concentration of 40%. After incubation at -20°C for 2 hours, proteins were precipitated by centrifugation for 15 min at 10,000×g. The pellet was freeze-dried and resuspended in 1×SDS-PAGE sample buffer (Bio-Rad, USA) and analyzed by 12% SDS-PAGE. Among the tested TFPs, TFP3 and TFP27 were identified to be effective for the secretion of human IL-32γ (FIG. 29).

## EXAMPLE 22

## TFP LIBRARY FROM PICHIA PASTORIS SELECTED IN SACCHAROMYCES CEREVISIAE

[0142] The TFP selection method of this invention could also be applied to other sources of the genomic or cDNA library. As an example of mRNA sources, the yeast *P. pastoris* was tested. Total RNA was isolated from yeast *P. pastoris* GS115 (Invitrogen, USA) for the construction of a cDNA library. Yeast was cultivated to mid-exponential phase in YPD media (2% yeast extract, 1% Bacto-peptone and 2% glucose). Total RNA was isolated from *P. pastoris* by a method described in Elion *et al.* (Elion *et al., Cell 39*:663 (1984)). Purification of Poly(A)$^+$ mRNA from total RNA was carried out using an Oligotex mRNA kit (Qiagen, Germany). cDNA was synthesized from the isolated mRNA using a SMART cDNA synthesis kit (BD Bioscience, USA). A specially designed primer ASA24N6 (SEQ ID NO:16) was used for the synthesis of the first strand cDNA instead of a primer included in the SMART kit as described in Example 4 (FIG. 8). Primer ASA24N6 could randomly bind to any position of mRNA due to its random hexameric sequence. Thus, most of first stranded cDNA amplified using this method contained the 5' partial sequence encoding the N-terminal part of yeast genes. The first stranded cDNA library with 5' partial sequence was used as a PCR template for double stranded cDNA synthesis with the 5' PCR

primer of the SMART Kit (BD Bioscience, USA) and primer ASA24 (SEQ ID NO:17). PCR products produced using this method contain numerous 5' partial fragments of cDNA with *Sfi*I sites at both ends. PCR conditions included one cycle of 95°C for 20 sec, and 20 cycles of 95°C for 30 sec, 68°C for 6 min as recommended in the kit. Amplified cDNA was treated with phenol/chloroform/isoamyl alcohol (25:24:1) and precipitated with 2 volumes of ethanol and 0.1 volume of 3 M sodium acetate (pH 5.0). Recovered cDNA was digested with *Sfi*I at 50°C for 2 hours and then fractionated using agarose gel electrophoresis. 0.5 to 1 kb DNA was isolated from the gel using a gel extraction kit (Bioneer, Korea). Extracted DNA was ligated into a *Sfi*I digested YGaINV vector and transformed into *E. coli* DH5α. Transformed *E. coli* was plated on LB media with ampicillin (1% Bacto-peptone, 0.5% yeast extract, 1% NaCl with 50 μg/ml ampicillin) and incubated at 37°C overnight. About $4\times10^4$ *E. coli* colonies were pooled with sterile distilled water and the total plasmids containing the cDNA library synthesized by random primer fused to the *SUC2* gene were isolated by using a plasmid isolation kit (Bioneer, Korea). For the selection of a TFP library secreting invertase from yeast *P. pastoris*, library DNA was transformed into *S. cerevisiae* Y2805 Δ*gal1*Δ*suc2*(*Mat a ura3 suc2::Tc190 pep4::HIS3 gal1 can1*) according to a lithium acetate method (Hill *et al.*, *Nucleic Acids Res. 19*:5791 (1991)). Transformed cells were spread on both UD media (0.67% yeast nitrogen base without amino acids, 0.77 g/l amino acid mixture, 2% glucose and 2% agar) and YPSGA media (1% yeast extract, 2% Bacto-peptone, 2% sucrose, 0.3% galactose, 1 μg/ml antimycin A, and 2% agar) and incubated at 30°C for 4 to 6 days. Around 1,000 transformants were obtained from the cDNA library of *P. pastoris*. Five different transformants grown on YPSGA media were randomly selected and total DNA was isolated from cultured cells of each colony using glass beads. Then the DNA was precipitated with ethanol. Isolated DNA was retransformed into *E. coli* DH5α. *E. coli* was plated on LB media with ampicillin (1% Bacto-peptone, 0.5% yeast extract, 1% NaCl with

50 μg/ml ampicillin) and incubated at 37°C overnight. Plasmids were isolated from transformed *E. coli* using a plasmid isolation kit (Bioneer, Korea). To analyze the sequence of each TFP obtained from the cDNA of *P. pastoris*, a sequencing primer GAL100-F (SEQ ID NO:12) binding to the *GAL10* promoter was used. Nucleotide sequences were determined by Genotech Co. (Taejon, Korea) using an automated sequencing unit (ABI Prism 377; PE Biosystems, Foster City, CA, USA). The sequences were tested on a BLAST search of the National Center for Biotechnology Information (NCBI) sequence database (www.ncbi.nlm.nih.gov). As a result, four different TFPs of *P. pastoris* were identified from plasmids isolated from 5 selected strains. The isolated plasmids were named pYHTS-PpTFP1, pYHTS-PpTFP2, pYHTS-PpTFP3, and pYHTS-PpTFP4. The four TFPs isolated from *P. pastoris* are summarized in Table 6.

Table 6. Isolated TFPs from *Pichia pastoris*

| Number of TFP | Homologue | Number of fused amino acids(signal) | Signal sequence | SEQ ID for protein | SEQ ID for DNA |
|---|---|---|---|---|---|
| PpTFP-1 | SUN family | 101 | Pre(21aa) | 84 | 85 |
| PpTFP-2 | SED1 | 94 | Pre(17aa) | 86 | 87 |
| PpTFP-3 | Unknown | 82 | Pre(20aa) | 88 | 89 |
| PpTFP-4 | Mucin-like | 127 | Pre(18aa) | 90 | 91 |

## EXAMPLE 23

## EVALUATION OF TFPS FROM *PICHIA PASTORIS* USING HUMAN IL2

[0143] Four *Pichia pastoris* TFPs summarized in Table 6 were tested for their secretion efficiency in *S. cerevisiae* using human IL-2. Each PpTFP was PCR amplified using primer pairs, PpTFP1-F (SEQ ID NO:227) and PpTFP1-R (SEQ ID NO:228), PpTFP2-F (SEQ ID NO:229) and PpTFP2-R (SEQ ID NO:230), PpTFP3-F (SEQ ID NO:231) and PpTFP3-R (SEQ ID NO:232), PpTFP4-F(SEQ ID NO:233) and PpTFP4-R (SEQ ID NO:234) from plasmids,

pYHTS-PpTFP1, pYHTS-PpTFP2, pYHTS-PpTFP3, and pYHTS-PpTFP4, respectively. Gel-purified PCR fragments were digested with *Sfi*I and subcloned into *Sfi*I digested YGaSW vector (FIG. 10) and the resulting plasmids were named as YGaSW-PpTFP1, YGaSW-PpTFP2, YGaSW-PpTFP3, and YGaSW-PpTFP4, respectively.

[0144] The amplified PCR fragment containing human IL-2 gene harboring homologous sequences with YGaSW-PpTFP vectors, was 2:1 mixed with *Swa*I digested YGaSW-PpTFP vectors and transformed into *S. cerevisiae* Y2805 (*Mat a ura3 SUC2 pep4::HIS3 GAL1 can1*) through *in vivo* recombination. Transforming cells were plated on UD media (0.67% yeast nitrogen base without amino acid, 0.77 g/l amino acid mixture, 2% glucose and 2% agar) and incubated at 30°C for 3 days. A single colony of each transformation was inoculated into YPDG broth (1% yeast extract, 2% bacto-peptone, 1% glucose and 1% galactose) and cultivated at 30°C for 40 hours. Culture supernatant (0.6 ml) was mixed with cold acetone for a final acetone concentration of 40%. After incubation at -20°C for 2 hours, proteins were precipitated by centrifugation for 15 min at 10,000×g. The pellet was freeze-dried and resuspended on 1×SDS-PAGE sample buffer (Bio-Rad, USA) and analyzed in 12% of SDS-PAGE. As shown in FIG. 30, all PpTFPs secreted human interleukin-2 into culture supernatant, suggesting the compatibility of TFP between two yeasts.

[0145] Having now fully described the invention, it will be understood by those of skill in the art that the same can be performed within a wide and equivalent range of conditions, formulations, and other parameters without affecting the scope of the invention or any embodiment thereof. All patents, patent applications and publications cited herein are fully incorporated by reference herein in their entirety.

| Applicant's or agent's file reference PCT/LP | International application No. TBA |
|---|---|

## INDICATIONS RELATING TO DEPOSITED MICROORGANISM OR OTHER BIOLOGICAL MATERIAL

(PCT Rule 13*bis*)

A. The indications made below relate to the deposited microorganism or other biological material referred to in the description on page 50, line 3

B. IDENTIFICATION OF DEPOSIT — Further deposits are identified on an additional sheet [X]

Name of depositary institution

Korean Collection for Type Cultures (KCTC)

Address of depositary institution *(including postal code and country)*

52, Oun-dong
Yusong-Ku
Taejon 305-333
Republic of Korea

| Date of deposit | Accession Number |
|---|---|
| July 14, 2005 | KCTC 10829BP |

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)* — This information is continued on an additional sheet [ ]

Escherichia coli
DH5@/pYGT17-IL2

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later *(specify the general nature of the indications e.g., "Accession Number of Deposit")*

| For receiving Office use only | For International Bureau use only |
|---|---|
| [ ] This sheet was received with the international application | [ ] This sheet was received by the International Bureau on: |
| Authorized officer | Authorized officer |

**Escherichia coli**
**DH5@/pYGT17-IL2**

**AUSTRALIA**

The applicant hereby gives notice that the furnishing of a sample of a microorganism shall only be effected prior to the grant of a patent, or prior to the lapsing, refusal or withdrawal of the application, to a person who is a skilled addressee without an interest in the invention (Regulation 3.25(3) of the Australian Patents Regulations).

**CANADA**

The applicant hereby requests that, until either a Canadian patent has been issued on the basis of the application or the application has been refused, or is abandoned and no longer subject to reinstatement, or is withdrawn, the furnishing of a sample of deposited biological material referred to in the application only be effected to an independent expert nominated by the Commissioner of Patents.

**DENMARK**

The applicant hereby requests that, until the application has been laid open to public inspection (by the Danish Patent Office), or has been finally decided upon by the Danish Patent Office without having been laid open to public inspection, the furnishing of a microorganism sample shall only be effected to an expert in the art (Sections 22 and 33(3) of the Danish Patents Act).

**FINLAND**

The applicant hereby requests that, until the publication of the mention of the grant of a patent by the National Board of Patents and Registration or for 20 years from the date of filing if the application has been finally decided upon without resulting in the grant of a patent by the National Board of Patents and Registration, the furnishing of a microorganism sample shall only be effected to an expert in the art.

**ICELAND**

The applicant hereby requests that, until a patent has been granted or a final decision taken by the Icelandic Patent Office concerning an application which has not resulted in a patent, the furnishing of a microorganism sample shall only be effected to an expert in the art (Sections 22 and 33(3) of the Icelandic Patent Act).

**NETHERLANDS**

The applicant hereby requests that until the date of a grant of a Netherlands patent or until the date on which the application is refused or withdrawn or lapsed, the microorganism shall be

made available as provided in Rule 31F(1) of the Patent Rules only by the issue of a sample to an expert.

**Escherichia coli**
**DH5@/pYGT17-IL2**

**NORWAY**

The applicant hereby requests that, until the application has been laid open to public inspection (by the Norwegian Patent Office), or has been finally decided upon by the Norwegian Patent Office without having been laid open to public inspection, the furnishing of a microorganism sample shall only be effected to an expert in the art (Sections 22 and 33(3) of the Norwegian Patents Act).

**SINGAPORE**

The applicant hereby requests that the furnishing of a sample of a microorganism shall only be made available to an expert.

**SWEDEN**

The applicant hereby requests that, until the application has been laid open to public inspection (by the Swedish Patent Office), or has been finally decided upon by the Swedish Patent Office without having been laid open to public inspection, the furnishing of a microorganism sample shall only be effected to an expert in the art.

**UNITED KINGDOM**

The applicant hereby requests that the furnishing of a sample of a microorganism shall only be made available to an expert.

WHAT IS CLAIMED IS:

1. A method of identifying a target protein specific translational fusion partner (TFP), said method comprising:

(i) co-transforming a plurality of reporter protein deficient host cells with a plurality of linear vectors and a nucleic acid encoding a target protein to produce a plurality of transformed host cells,

wherein each of said linear vectors comprises a nucleic acid fragment from a library of nucleic acid fragments and a nucleotide sequence encoding a N-terminal amino acid-deleted reporter protein, and

wherein said nucleic acid encoding a target protein comprises, at the 3' end, a nucleotide sequence encoding the N-terminal amino acids deleted from said reporter protein in said linear vector, and at the 5' end, a linker DNA;

(ii) incubating said plurality of transformed host cells under conditions effective to allow *in vivo* recombination of said linear vectors and said nucleic acid encoding a target protein;

(iii) identifying a cell showing an activity of the reporter protein from the plurality of transformed host cells of (ii); and

(iv) identifying a TFP from the cell identified in (iii);

wherein said TFP comprises a nucleic acid fragment which induces the secretion of said target protein.

2. A method of identifying a target protein specific TFP library, said method comprising:

(i) co-transforming a plurality of reporter protein deficient host cells with a plurality of linear vectors and a nucleic acid encoding a target protein to produce a plurality of transformed host cells,

wherein each of said linear vectors comprises a nucleic acid fragment from a library of nucleic acid fragments and a nucleotide sequence encoding a N-terminal amino acid-deleted reporter protein, and

wherein said nucleic acid encoding a target protein comprises, at the 3' end, a nucleotide sequence encoding the N-terminal amino acids deleted from said reporter protein in said linear vector, and at the 5' end, a linker DNA;

(ii) incubating said plurality of transformed host cells under conditions effective to allow *in vivo* recombination of said linear vectors and said nucleic acid encoding a target protein;

(iii) identifying cells showing an activity of the reporter protein from the plurality of transformed host cells of (ii); and

(iv) identifying a TFP library from the cells identified in (iii);

wherein said TFP library comprises nucleic acid fragments which individually induce the secretion of said target protein.

3. The method of claim 1 or 2, wherein said library of nucleic acid fragments is from genomic DNA or cDNA of a plant, bacteria, yeast, fungus, or animal.

4. The method of claim 1 or 2, wherein said library of nucleic acid fragments is from recombinant DNA.

5. The method of claim 3, wherein said yeast is selected from *Candida*, *Debaryomyces*, *Hansenula*, *Kluyveromyces*, *Pichia*, *Schizosaccharomyces*, *Yarrowia*, *Saccharomyces*, *Schwanniomyces*, and *Arxula* species.

6. The method of claim 5, wherein said yeast is selected from *Candida utilis*, *Candida boidinii*, *Candida albicans*, *Kluyveromyces lactis*, *Pichia pastoris*, *Pichia stipitis*, *Schizosaccharomyces pombe*, *Saccharomyces cerevisiae*, *Hansenula polymorpha*, *Yarrowia lipolytica*, *Schwanniomyces occidentalis*, and *Arxula adeninivorans*.

7. The method of claim 3, wherein said fungus is selected from *Aspergillus*, *Penicillium*, *Rhizopus*, and *Trichoderma* species.

8. The method of claim 3, wherein said bacteria is selected from *Escherichia, Pseudomonas* and *Bacillus* species.

9. The method of claim 3, wherein said plant is selected from *Arabidopsis*, maize, tobacco, and potato.

10. The method of claim 3, wherein said animal is selected from human, mouse, rat, rabbit, dog, cat, and monkey.

11. The method of any of claims 1-10, wherein said library of nucleic acid fragments is a library of pre-selected candidate TFPs.

12. The method of claim 11, wherein said library of pre-selected candidate TFPs is obtained by transforming a plurality of reporter protein-deficient host cells with a variety of vectors comprising a library of nucleic acid fragments and a nucleic acid encoding a reporter protein, collecting cells that grow, isolating vectors from the cells, and isolating nucleic acid fragments from the vectors, thereby obtaining a TFP library comprising nucleic acid fragments which individually induce secretion of the reporter protein.

13. The method of claim 11, wherein said library of pre-selected candidate TFPs is derived from sequences identified in a genome database by searching for (i) genes containing a pre-secretion signal homologous with those of one or more previously identified TFPs; (ii) genes comprising a secretion signal sequence, or (iii) genes encoding proteins passing through the endoplasmic reticulum.

14. The method of claim 11, wherein said library of pre-selected candidate TFPs is obtained by diversifying previously identified TFPs.

15. The method of claim 11, wherein said library of pre-selected candidate TFPs is obtained by artificially designing nucleic acid fragments to have the pre and pro signal sequence swapped between previously identified TFPs.

16. The method of claim 11, wherein said library of pre-selected candidate TFPs is a library of core TFPs wherein the core TFPs are a collection of previously identified TFPs that are effective for one or more target proteins.

17. The method of any of claims 1-16, wherein said nucleic acid fragments have a size of fewer than 1000 base pairs.

18. The method of claim 17, wherein said nucleic acid fragments have a size of fewer than 700 base pairs.

19. The method of claim 18, wherein said nucleic acid fragments have a size of fewer than 500 base pairs.

20. The method of claim 19, wherein said nucleic acid fragments have a size of fewer than 300 base pairs.

21. The method of any of claims 1-20, wherein said library of nucleic acid fragments is constructed by enzymatic cleavage of the DNA.

22. The method of any of claims 1-20, wherein said library of nucleic acid fragments is constructed by cDNA synthesis.

23. The method of any of claims 1-20, wherein said library of nucleic acid fragments is constructed by recombinant DNA technology.

24. The method of claim 23, wherein said recombinant DNA technology comprises unidirectional deletion.

25. The method of any of claims 1-24, wherein said host cell is selected from plant, bacterial, fungal, yeast, or animal cells.

26. The method of claim 25, wherein said yeast is selected from *Candida*, *Debaryomyces*, *Hansenula*, *Kluyveromyces*, *Pichia*, *Schizosaccharomyces*, *Yarrowia*, *Saccharomyces*, *Schwanniomyces*, and *Arxula* species.

27. The method of claim 26, wherein said yeast is selected from *Candida utilis*, *Candida boidinii*, *Candida albicans*, *Kluyveromyces lactis*, *Pichia pastoris*, *Pichia stipitis*, *Schizosaccharomyces pombe*, *Saccharomyces cerevisiae*, *Hansenula polymorpha*, *Yarrowia lipolytica*, *Schwanniomyces occidentalis*, and *Arxula adeninivorans*.

28. The method of claim 25, wherein said fungus is selected from *Aspergillus*, *Penicillium*, *Rhizopus*, and *Trichoderma* species.

29. The method of claim 25, wherein said bacteria is selected from *Escherichia*, *Pseudomonas*, and *Bacillus* species.

30. The method of claim 25, wherein said plant is selected from *Arabidopsis*, maize, tobacco, and potato.

31. The method of claim 25, wherein said animal is selected from human, mouse, rat, rabbit, dog, cat, monkey, and insect.

32. The method of claim 25, wherein said animal cells are selected from CHO, COS 1, COS 7, BSC 1, BSC 40, BMT 10, and Sf9.

33. The method of any of claims 1-27, wherein said host cells are yeast cells, and the nucleic acid fragments are isolated from the genome or cDNA of a yeast.

34. The method of any of claims 1-33, wherein said reporter protein is a protein that is secreted into the extracellular space.

35. The method of claim 34, wherein said reporter protein is selected from invertase, sucrase, cellulase, xylanase, maltase, amylase, glucoamylase, galactosidase, phosphatase, beta-lactamase, lipase or protease.

36. The method of claim 35, wherein said galactosidase is selected from alpha-galactosidase, beta-galactosidase, and melibiase.

37. The method of claim 36, wherein said reporter protein is melibiase.

38. The method of claim 35, wherein said phosphatase is PHO5.

39. The method of claim 35, wherein said host cells are yeast, said reporter protein is invertase and the transformed yeast cells are selected for their ability to grow on sucrose or raffinose.

40. The method of claim 35, wherein said host cells are yeast, said reporter protein is amylase, the yeast cells are non-amylolytic, and the transformed cells are screened for their ability to degrade starch.

41. The method of any of claims 1-33, wherein said step of identifying cells showing an activity of the reporter protein occurs by using a reporter protein which provides resistance to a growth inhibitor.

42. The method of any of claims 1-41, wherein said step of identifying cells showing an activity of the reporter protein occurs by using two or more reporter proteins.

43. The method of claim 42, wherein said step of identifying cells showing an activity of the reporter protein occurs by using two reporter proteins.

44. The method of claim 43, wherein said two reporter proteins are lipase and invertase.

45. The method of any of claims 1-44, wherein said target protein is from a plant, animal, or microorganism.

46. The method of claim 45, wherein said target protein is a human protein.

47. The method of claim 45, wherein said target protein is a cytokine, serum protein, colony stimulating factor, growth factor, hormone, or enzyme.

48. The method of claim 45, wherein said target protein is selected from an interleukin, coagulation factor, interferon-α, -β or -γ, granulocyte-colony stimulating factor, human granulocyte macrophage-colony stimulating factor, tissue growth factors, epithelial growth factors, TGFα, TGFβ, epidermal growth factor, platelet-derived growth factor, fibroblast growth factor, follicle stimulating hormone, thyroid stimulating hormone, antidiuretic hormone, pigmentary hormone, parathyroid hormone, luteinizing hormone-releasing hormone, carbohydrate-specific enzymes, proteolytic enzymes, lipases, oxidoreductases, transferases, hydrolases, lyases, isomerases, ligases, immunoglobulins, cytokine receptors, lactoferrin, phospholipase A2-activating protein, insulin, tumor necrosis factor, calcitonin, calcitonin gene related peptide, enkephalin, somatomedin, erythropoietin, hypothalamic releasing

factor, prolactin, chorionic gonadotropin, tissue plasminogen activator, growth hormone releasing peptide, thymic humoral factor, anticancer peptides, or antibiotic peptides.

49. The method of claim 45, wherein said target protein is selected from human interleukin-2, human interleukin-1β, human interleukin-6, human interleukin-32α, -32β or -32γ, Factor VII, Factor VIII, Factor IX, human serum albumin, human interferon-α, -β or -γ, human granulocyte-colony stimulating factor, human granulocyte macrophage-colony stimulating factor, human growth hormone, human platelet-derived growth factor, human basic fibroblast growth factor, human epidermal growth factor, human insulin-like growth factor, human nerve growth factor, human transforming growth factor β–1, human follicle stimulating hormone, glucose oxidase, glucodase, galactosidase, glucocerebrosidase, glucuronidase, asparaginase, arginase, arginine deaminase, peroxide dismutase, endotoxinase, catalase, chymotrypsin, uricase, adenosine diphosphatase, tyrosinase, bilirubin oxidase, bovine galactose-1-phosphate uridyltransferase, jellyfish green fluorescent protein, *Candida antarctica* lipase B, *Candida rugosa* lipase, fungal chloroperoxidase, β-galactosidase, resolvase, α-galactosidase, β-glucosidase, trehalose synthase, cyclodextrin glycosyl transferase, xylanase, phytase, human lactoferrin, human erythropoietin, human paraoxonase, human growth differentiation factor 15, human galectin-3 binding protein, human serine protease inhibitor, Kunitz type 2, human Janus kinase 2, human fms-like tyrosine kinase 3 ligand, human YM1 & 2, human CEMI, human diacylglycerol acyltransferase, human leptin, human mL259, human proteinase 3, human lysozyme, human DEAD box protein 41, human etoposide induced protein 24, mouse caspase1, bovine angiogenin, and earthworm lumbrokinase.

50. The method of claim 45, wherein said target protein is a protein that is difficult to produce using conventional recombinant production methods.

51. The method of any of claims 1-50, wherein said linker DNA is more than 20 base pairs in length.

52. The method of claim 51, wherein said linker DNA is more than 30 base pairs in length.

53. The method of claim 52, wherein said linker DNA is more than 40 base pairs in length.

54. The method of any of claims 1-53, wherein said linker DNA encodes a protease recognition sequence thereby allowing cleavage at the junction of the TFP and the target protein.

55. The method of claim 54, wherein said linker DNA encodes a yeast kex2p-recognition sequence.

56. The method of claim 55, wherein said linker DNA encodes an amino acid sequence comprising Lys-Arg or Arg-Arg.

57. The method of claim 56, wherein said linker DNA encodes an amino acid sequence comprising Leu-Asp-Lys-Arg (SEQ ID NO:214).

58. The method of claim 54, wherein said linker DNA encodes a mammalian furin-recognition sequence.

59. The method of claim 58, wherein said linker DNA comprises encodes an amino acid sequence comprising Arg-X-X-Arg.

60. The method of claim 54, wherein said linker DNA encodes a factor Xa-recognition sequence.

61. The method of claim 59, wherein said linker DNA encodes an amino acid sequence comprising Ile-Glu-Gly-Arg (SEQ ID NO:215).

62. The method of claim 54, wherein said linker DNA encodes an enterokinase-recognition sequence.

63. The method of claim 62, wherein said linker DNA encodes an amino acid sequence comprising Asp-Asp-Lys.

64. The method of claim 54, wherein said linker DNA encodes a subtilisin-recognition sequence.

65. The method of claim 64, wherein said linker DNA encodes an amino acid sequence comprising Ala-Ala-His-Tyr (SEQ ID NO:216).

66. The method of claim 54, wherein said linker DNA encodes a tobacco etch virus protease-recognition sequence.

67. The method of claim 66, wherein said linker DNA encodes an amino acid sequence comprising Glu-Asn-Leu-Tyr-Phe-Gln-Gly (SEQ ID NO:217).

68. The method of claim 54, wherein said linker DNA encodes a thrombin-recognition sequence.

69. The method of claim 68, wherein said linker DNA encodes an amino acid sequence comprising Arg-Gly-Pro-Arg (SEQ ID NO:218).

70. The method of claim 54, wherein said linker DNA encodes a ubiquitin hydrolase-recognition sequence.

71. The method of claim 70, wherein said linker DNA encodes an amino acid sequence comprising Arg-Gly-Gly.

72. The method of any of claims 1-71, wherein said linker DNA encodes an affinity tag.

73. The method of claim 72, wherein said affinity tag is selected from GST, MBP, NusA, thioredoxin, ubiquitin, FLAG, BAP, 6HIS, STREP, CBP, CBD, and S-tag.

74. The method of any of claims 1-73, wherein said linker DNA encodes a restriction enzyme recognition site.

75. The method of claim 74, wherein said restriction enzyme recognition site is for SfiI.

76. The method of claim 75, wherein said linker DNA further encodes kex2p-like protease- or kex2p-recognition sequence.

77. A TFP identified by the method of any of claims 1-76 or a fragment or derivative thereof.

78. The TFP of claim 77, wherein said TFP is selected from the group consisting of TFP-9 (SEQ ID NO:29), TFP-13 (SEQ ID NO:31), TFP-17 (SEQ ID NO:33), TFP-18 (SEQ ID NO:35), TFP-19 (SEQ ID NO:37), TFP-20 (SEQ ID NO:39), TFP-21 (SEQ ID NO:41), TFP-25 (SEQ ID NO:43), TFP-27 (SEQ ID NO:45), TFP-11 (SEQ ID NO:61), TFP-22 (SEQ ID NO:63), TFP-29 (SEQ ID NO:65), TFP-34 (SEQ ID NO:67), TFP-38 (SEQ ID NO:69), TFP-39 (SEQ ID NO:129), TFP-43 (SEQ ID NO:131), TFP-44 (SEQ ID NO:133), TFP-48 (SEQ ID NO:135), TFP-52 (SEQ ID NO:137), TFP-54 (SEQ ID NO:139), TFP-40 (SEQ ID NO:175), TFP-50 (SEQ ID

NO:177), TFP-51 (SEQ ID NO:179), TFP-57 (SEQ ID NO:181), TFP-58 (SEQ ID NO:183), TFP-59 (SEQ ID NO:185), TFP-5 (SEQ ID NO:200), TFP-6 (SEQ ID NO:202), TFP-7 (SEQ ID NO:204), TFP-8 (SEQ ID NO:206), PpTFP-1 (SEQ ID NO:84), PpTFP-2 (SEQ ID NO:86), PpTFP-3 (SEQ ID NO:88), and PpTFP-4 (SEQ ID NO:90) or a fragment or derivative thereof.

79. A TFP library comprising one or more TFPs identified by the method of any of claims 1-76 or a fragment or derivative thereof.

80. The TFP library of claim 76, comprising two or more TFPs selected from the group consisting of TFP-9 (SEQ ID NO:29), TFP-13 (SEQ ID NO:31), TFP-17 (SEQ ID NO:33), TFP-18 (SEQ ID NO:35), TFP-19 (SEQ ID NO:37), TFP-20 (SEQ ID NO:39), TFP-21 (SEQ ID NO:41), TFP-25 (SEQ ID NO:43), TFP-27 (SEQ ID NO:45), TFP-11 (SEQ ID NO:61), TFP-22 (SEQ ID NO:63), TFP-29 (SEQ ID NO:65), TFP-34 (SEQ ID NO:67), TFP-38 (SEQ ID NO:69), TFP-39 (SEQ ID NO:129), TFP-43 (SEQ ID NO:131), TFP-44 (SEQ ID NO:133), TFP-48 (SEQ ID NO:135), TFP-52 (SEQ ID NO:137), TFP-54 (SEQ ID NO:139), TFP-40 (SEQ ID NO:175), TFP-50 (SEQ ID NO:177), TFP-51 (SEQ ID NO:179), TFP-57 (SEQ ID NO:181), TFP-58 (SEQ ID NO:183), TFP-59 (SEQ ID NO:185), TFP-5 (SEQ ID NO:200), TFP-6 (SEQ ID NO:202), TFP-7 (SEQ ID NO:204), TFP-8 (SEQ ID NO:206), PpTFP-1 (SEQ ID NO:84), PpTFP-2 (SEQ ID NO:86), PpTFP-3 (SEQ ID NO:88), and PpTFP-4 (SEQ ID NO:90) or a fragment or derivative thereof.

81. The TFP library of claim 80, comprising four or more of the listed TFPs or a fragment or derivative thereof.

82. The TFP library of claim 81, comprising six or more of the listed TFPs or a fragment or derivative thereof.

83. The TFP library of claim 82, comprising eight or more of the listed TFPs or a fragment or derivative thereof.

84. The TFP library of claim 83, comprising ten or more of the listed TFPs or a fragment or derivative thereof.

85. The TFP library of claim 84, comprising twelve or more of the listed TFPs or a fragment or derivative thereof.

86. The TFP library of claim 79, comprising six or more TFPs selected from the group consisting of TFP-9 (SEQ ID NO:29), TFP-13 (SEQ ID NO:31), TFP-17 (SEQ ID NO:33), TFP-18 (SEQ ID NO:35), TFP-19 (SEQ ID NO:37), TFP-20 (SEQ ID NO:39), TFP-21 (SEQ ID NO:41), TFP-25 (SEQ ID NO:43), TFP-27 (SEQ ID NO:45), TFP-11 (SEQ ID NO:61), TFP-22 (SEQ ID NO:63), TFP-29 (SEQ ID NO:65), TFP-34 (SEQ ID NO:67), TFP-38 (SEQ ID NO:69), TFP-39 (SEQ ID NO:129), TFP-43 (SEQ ID NO:131), TFP-44 (SEQ ID NO:133), TFP-48 (SEQ ID NO:135), TFP-52 (SEQ ID NO:137), TFP-54 (SEQ ID NO:139), TFP-40 (SEQ ID NO:175), TFP-50 (SEQ ID NO:177), TFP-51 (SEQ ID NO:179), TFP-57 (SEQ ID NO:181), TFP-58 (SEQ ID NO:183), TFP-59 (SEQ ID NO:185), TFP-5 (SEQ ID NO:200), TFP-6 (SEQ ID NO:202), TFP-7 (SEQ ID NO:204), TFP-8 (SEQ ID NO:206), PpTFP-1 (SEQ ID NO:84), PpTFP-2 (SEQ ID NO:86), PpTFP-3 (SEQ ID NO:88), PpTFP-4 (SEQ ID NO:90), TFP-1 (SEQ ID NO:219), TFP-2 (SEQ ID NO:221), TFP-3 (SEQ ID NO:223), TFP-4 (SEQ ID NO:225), and TFP 32 (SEQ ID NO:208) or a fragment or derivative thereof.

87. The TFP library of claim 86, comprising eight or more of the listed TFPs or a fragment or derivative thereof.

88. The TFP library of claim 87, comprising ten or more of the listed TFPs or a fragment or derivative thereof.

89. The TFP library of claim 88, comprising twelve or more of the listed TFPs or a fragment or derivative thereof.

90. The TFP library of claim 89, comprising fifteen or more of the listed TFPs or a fragment or derivative thereof.

91. A library of nucleic acid fragments, comprising 10 or more nucleic acid fragments identified by the method of claim 12.

92. The library of nucleic acid fragments of claim 91, comprising 50 or more nucleic acid fragments identified by the method of claim 12.

93. The library of nucleic acid fragments of claim 92, comprising 100 or more nucleic acid fragments identified by the method of claim 12.

94. The library of nucleic acid fragments of claim 93, comprising 500 or more nucleic acid fragments identified by the method of claim 12.

95. The library of nucleic acid fragments of claim 94, comprising 1000 or more nucleic acid fragments identified by the method of claim 12.

96. The library of nucleic acid fragments of claim 95, comprising 2000 or more nucleic acid fragments identified by the method of claim 12.

97. A library of nucleic acid fragments, comprising 10 or more nucleic acid fragments identified by the method of claim 13.

98. The library of nucleic acid fragments of claim 97, comprising 50 or more nucleic acid fragments identified by the method of claim 13.

99. The library of nucleic acid fragments of claim 98, comprising 100 or more nucleic acid fragments identified by the method of claim 13.

100. A library of nucleic acid fragments, comprising 10 or more nucleic acid fragments identified by the method of claim 14.

101. The library of nucleic acid fragments of claim 100, comprising 50 or more nucleic acid fragments identified by the method of claim 14.

102. The library of nucleic acid fragments of claim 101, comprising 100 or more nucleic acid fragments identified by the method of claim 14.

103. The library of nucleic acid fragments of claim 102, comprising 500 or more nucleic acid fragments identified by the method of claim 14.

104. The library of nucleic acid fragments of claim 103, comprising 1000 or more nucleic acid fragments identified by the method of claim 14.

105. A library of nucleic acid fragments, comprising 10 or more nucleic acid fragments identified by the method of claim 15.

106. The library of nucleic acid fragments of claim 105, comprising 50 or more nucleic acid fragments identified by the method of claim 15.

107. The library of nucleic acid fragments of claim 106, comprising 100 or more nucleic acid fragments identified by the method of claim 15.

108. The library of nucleic acid fragments of claim 107, comprising 500 or more nucleic acid fragments identified by the method of claim 15.

109. A nucleic acid comprising a nucleotide sequence encoding a TFP or a fragment or derivative thereof and a nucleic acid sequence encoding a target protein, wherein said TFP is selected from the group consisting of TFP-9 (SEQ ID NO:29), TFP-13 (SEQ ID NO:31), TFP-17 (SEQ ID NO:33), TFP-18 (SEQ ID NO:35), TFP-19 (SEQ ID NO:37), TFP-20 (SEQ ID NO:39), TFP-21 (SEQ ID NO:41), TFP-25 (SEQ ID NO:43), TFP-27 (SEQ ID NO:45), TFP-11 (SEQ ID NO:61), TFP-22 (SEQ ID NO:63), TFP-29 (SEQ ID NO:65), TFP-34 (SEQ ID NO:67), TFP-38 (SEQ ID NO:69), TFP-39 (SEQ ID NO:129), TFP-43 (SEQ ID NO:131), TFP-44 (SEQ ID NO:133), TFP-48 (SEQ ID NO:135), TFP-52 (SEQ ID NO:137), TFP-54 (SEQ ID NO:139), TFP-40 (SEQ ID NO:175), TFP-50 (SEQ ID NO:177), TFP-51 (SEQ ID NO:179), TFP-57 (SEQ ID NO:181), TFP-58 (SEQ ID NO:183), TFP-59 (SEQ ID NO:185), TFP-5 (SEQ ID NO:200), TFP-6 (SEQ ID NO:202), TFP-7 (SEQ ID NO:204), TFP-8 (SEQ ID NO:206), PpTFP-1 (SEQ ID NO:84), PpTFP-2 (SEQ ID NO:86), PpTFP-3 (SEQ ID NO:88), and PpTFP-4 (SEQ ID NO:90).

110. The nucleic acid of claim 109, wherein said target protein is selected from IL-2, IL-32, human growth hormone and human caspase-1 subunit P10.

111. The nucleic acid of claim 109, wherein said TFP is selected from the group consisting of TFP-9 (SEQ ID NO:29), TFP-13 (SEQ ID NO:31), TFP-17 (SEQ ID NO:33), TFP-18 (SEQ ID NO:35), TFP-19 (SEQ ID NO:37), TFP-20 (SEQ ID NO:39), TFP-21 (SEQ ID NO:41), TFP-25 (SEQ ID NO:43), TFP-27 (SEQ ID NO:45), PpTFP-1 (SEQ ID NO:84), PpTFP-2 (SEQ ID NO:86), PpTFP-3 (SEQ ID NO:88), PpTFP-4 (SEQ ID NO:90) and said target protein is IL-2.

112. The nucleic acid of claim 109, wherein said TFP is selected from the group consisting of TFP-11 (SEQ ID NO:61), TFP-22 (SEQ ID NO:63), TFP-29 (SEQ ID NO:65), TFP-34 (SEQ ID NO:67), TFP-38 (SEQ ID NO:69), and said target protein is IL-32 alpha.

113. The nucleic acid of claim 109, wherein said TFP is selected from the group consisting of TFP-9 (SEQ ID NO:29), TFP-13 (SEQ ID NO:31), TFP-17 (SEQ ID NO:33), TFP-18 (SEQ ID NO:35), TFP-19 (SEQ ID NO:37), TFP-20 (SEQ ID NO:39), TFP-21 (SEQ ID NO:41), TFP-25 (SEQ ID NO:43), TFP-27 (SEQ ID NO:45), TFP-11 (SEQ ID NO:61), TFP-22 (SEQ ID NO:63), TFP-29 (SEQ ID NO:65), TFP-34 (SEQ ID NO:67), TFP-38 (SEQ ID NO:69), and said target protein is growth hormone.

114. A method of producing a target protein, comprising preparing a vector comprising a nucleotide sequence encoding said target protein operably linked to a nucleotide sequence encoding a TFP identified by the method of any of claims 1-76 or a fragment or derivative thereof, transforming a host cell with said vector, and culturing said host cell under conditions in which the target protein is produced.

115. The method of claim 114, wherein said vector comprises the nucleic acid of claim 107.

116. The method of claim 115, wherein said target protein is selected from IL-2, IL-32, human growth hormone and human caspase-1 subunit P10.

117. A linear vector comprising a nucleic acid fragment from a library of nucleic acid fragments and a nucleotide sequence encoding a N-terminal amino acid-deleted reporter protein.

118. The linear vector of claim 117, further comprising a nucleotide sequence encoding a target protein.

119. A plurality of reporter protein-deficient host cells transformed with a plurality of linear vectors and a nucleic acid encoding a target protein,

wherein each of said linear vectors comprises a nucleic acid fragment from a library of nucleic acid fragments and a nucleotide sequence encoding a N-terminal amino acid-deleted reporter protein, and

wherein said nucleic acid encoding a target protein comprises, at the 3' end, a nucleotide sequence encoding the N-terminal amino acids deleted from said reporter protein in said linear vector, and at the 5' end, a linker DNA.

FIG. 1

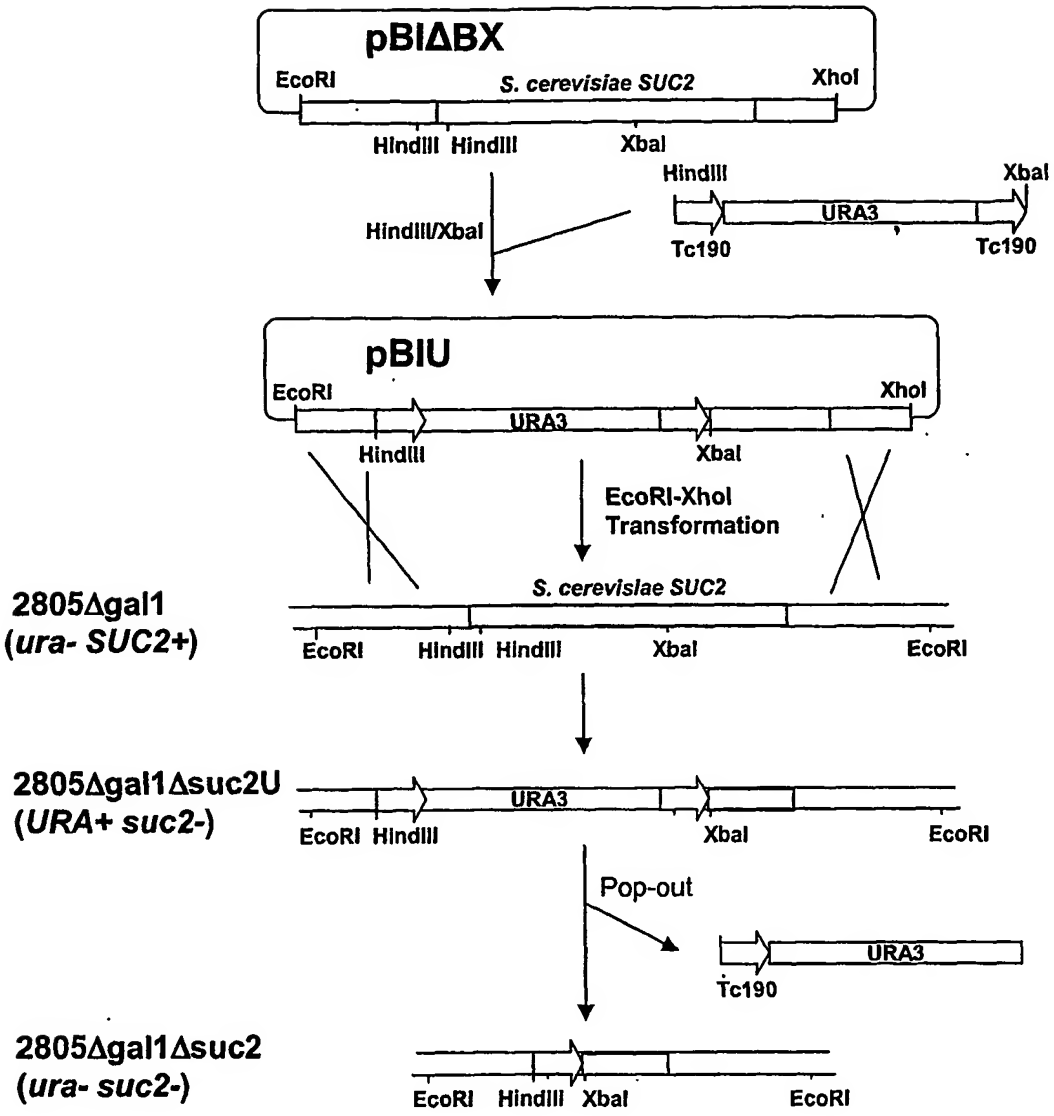

pBIΔBX
EcoRI
S. cerevisiae SUC2
XhoI
HindIII HindIII
XbaI
HindIII/XbaI
HindIII
XbaI
URA3
Tc190
Tc190
pBIU
EcoRI
XhoI
URA3
HindIII
XbaI
EcoRI-XhoI
Transformation
2805Δgal1
(ura- SUC2+)
S. cerevisiae SUC2
EcoRI
HindIII HindIII
XbaI
EcoRI
2805Δgal1Δsuc2U
(URA+ suc2-)
URA3
EcoRI HindIII
XbaI
EcoRI
Pop-out
URA3
Tc190
2805Δgal1Δsuc2
(ura- suc2-)
EcoRI
HindIII
XbaI
EcoRI

FIG. 2

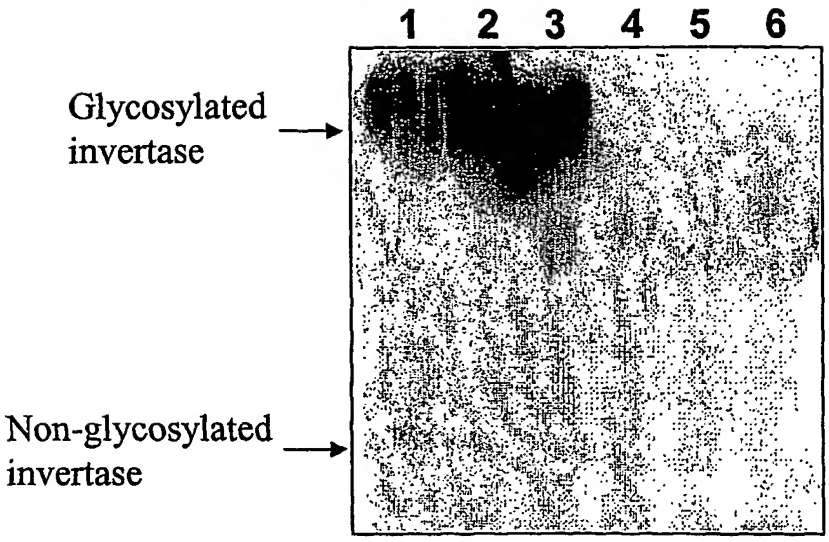

1 2 3 4 5 6
Glycosylated invertase
Non-glycosylated invertase

FIG. 3

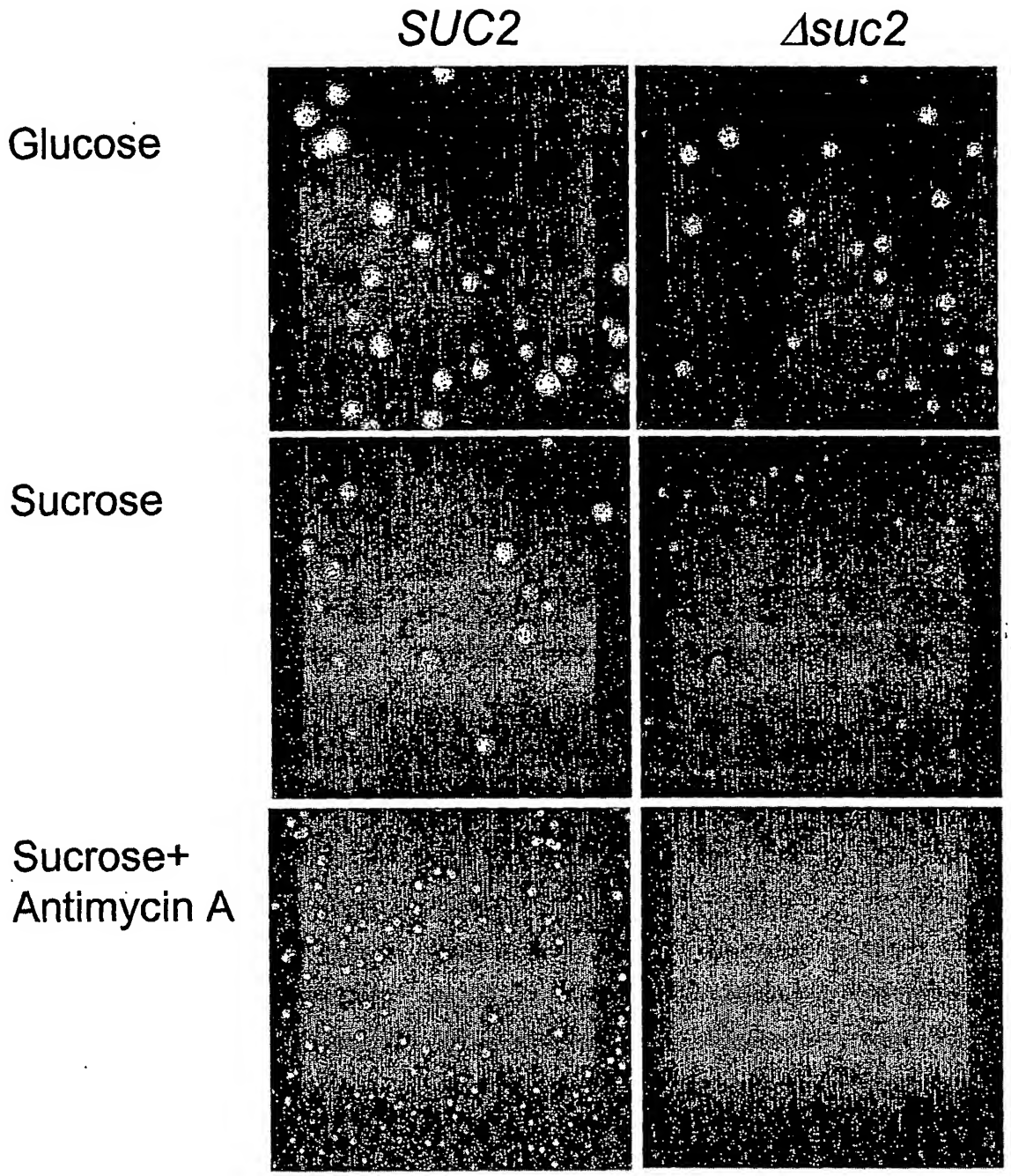


FIG. 4

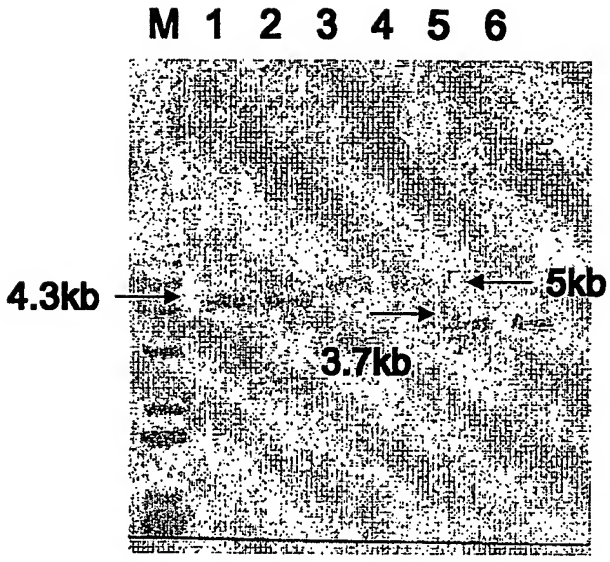


FIG. 5

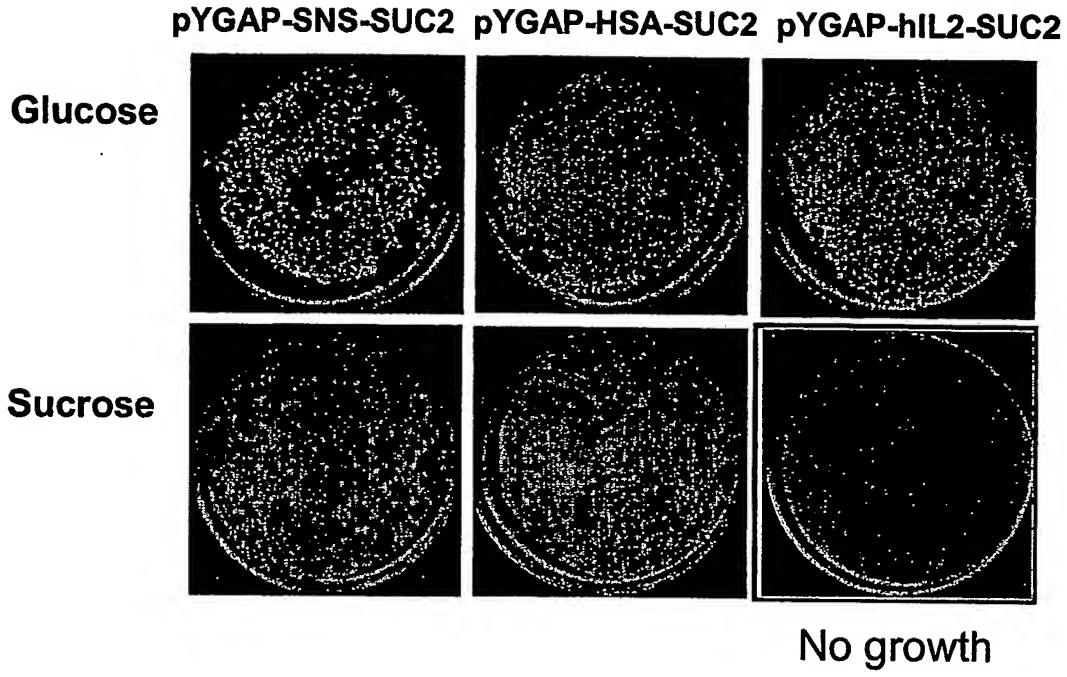

pYGAP-SNS-SUC2
pYGAP-HSA-SUC2
pYGAP-hIL2-SUC2
Glucose
Sucrose
No growth

FIG. 6

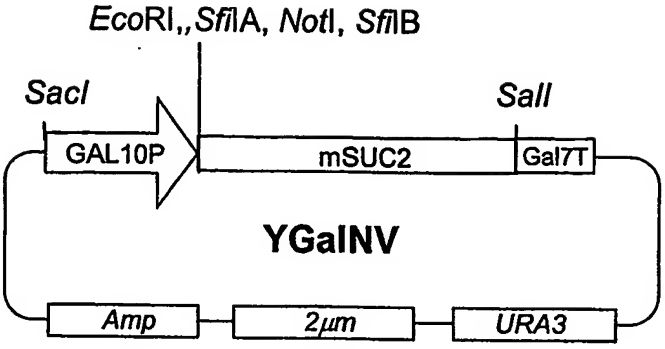

EcoRI,,SfiIA, NotI, SfiIB
SacI
SalI
GAL10P
mSUC2
Gal7T
YGalNV
Amp
2μm
URA3

FIG. 7

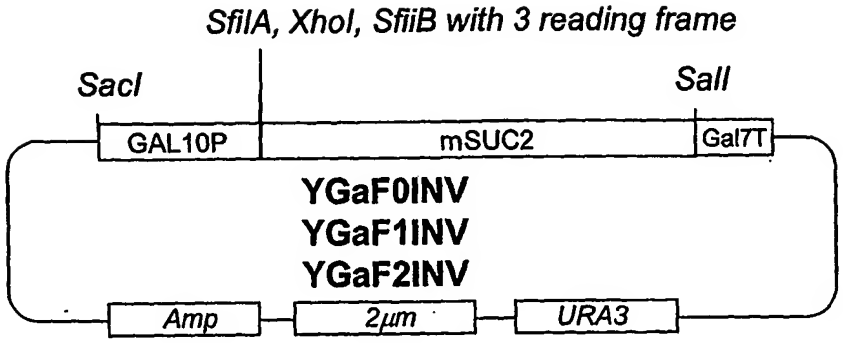

SfilA, XhoI, SfiiB with 3 reading frame
SacI
SalI
GAL10P
mSUC2
Gal7T
YGaF0INV
YGaF1INV
YGaF2INV
Amp
2μm
URA3

FIG. 8

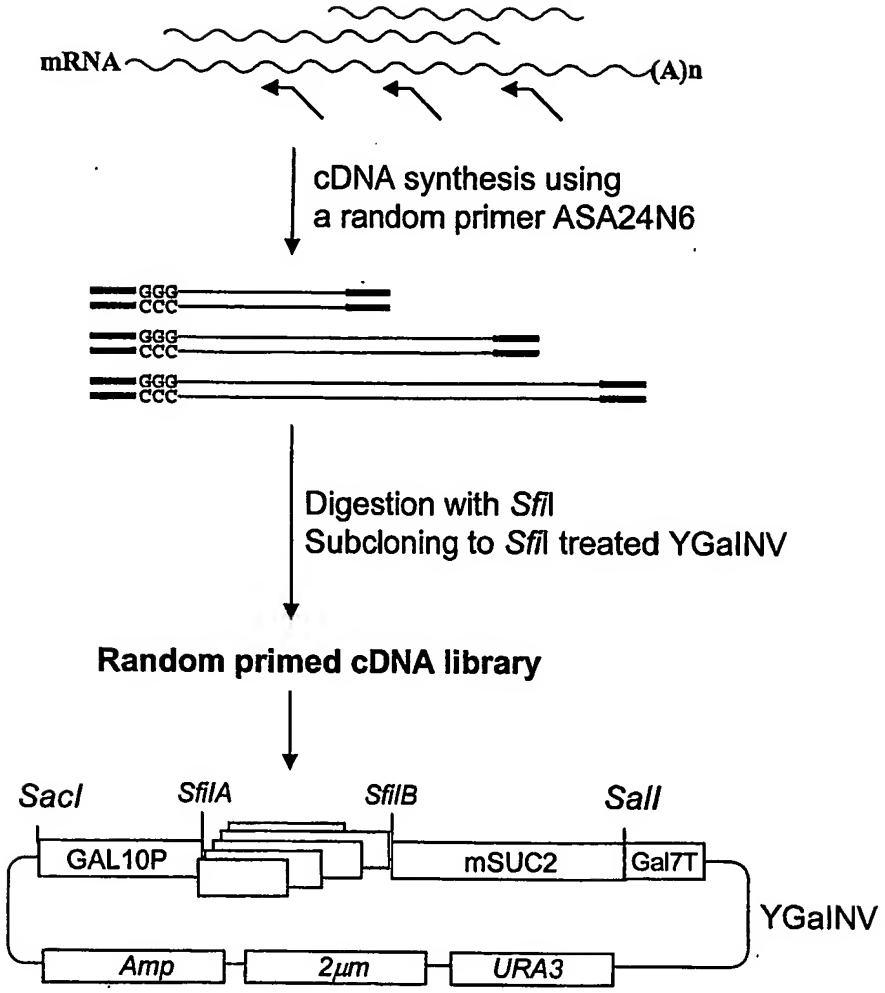

mRNA
(A)n
cDNA synthesis using
a random primer ASA24N6
GGG
CCC
Digestion with SfiI
Subcloning to SfiI treated YGalNV
Random primed cDNA library
SacI
SfiIA
SfiIB
SalI
GAL10P
mSUC2
Gal7T
YGalNV
Amp
2µm
URA3

FIG. 9

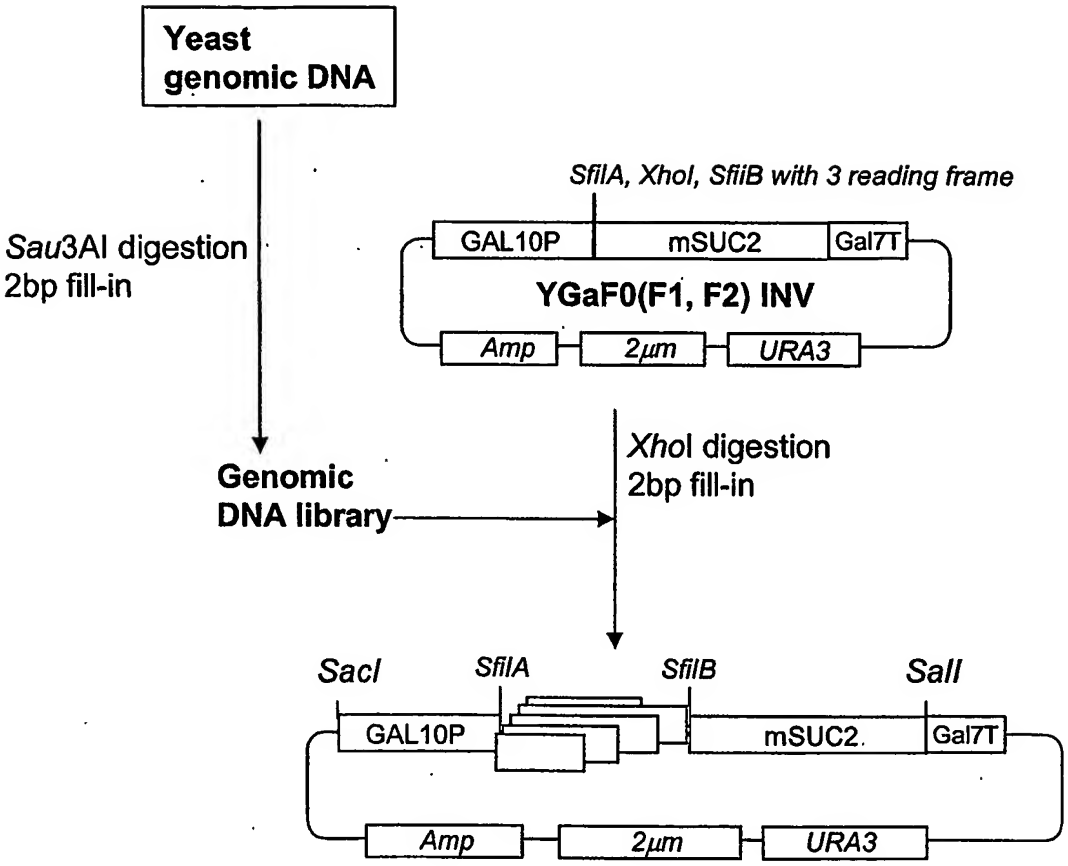

Yeast genomic DNA
Sau3AI digestion
2bp fill-in
Genomic DNA library
SfiIA, XhoI, SfiiB with 3 reading frame
GAL10P
mSUC2
Gal7T
YGaF0(F1, F2) INV
Amp
2µm
URA3
XhoI digestion
2bp fill-in
SacI
SfiIA
SfiIB
SalI
GAL10P
mSUC2
Gal7T
Amp
2µm
URA3

FIG. 10

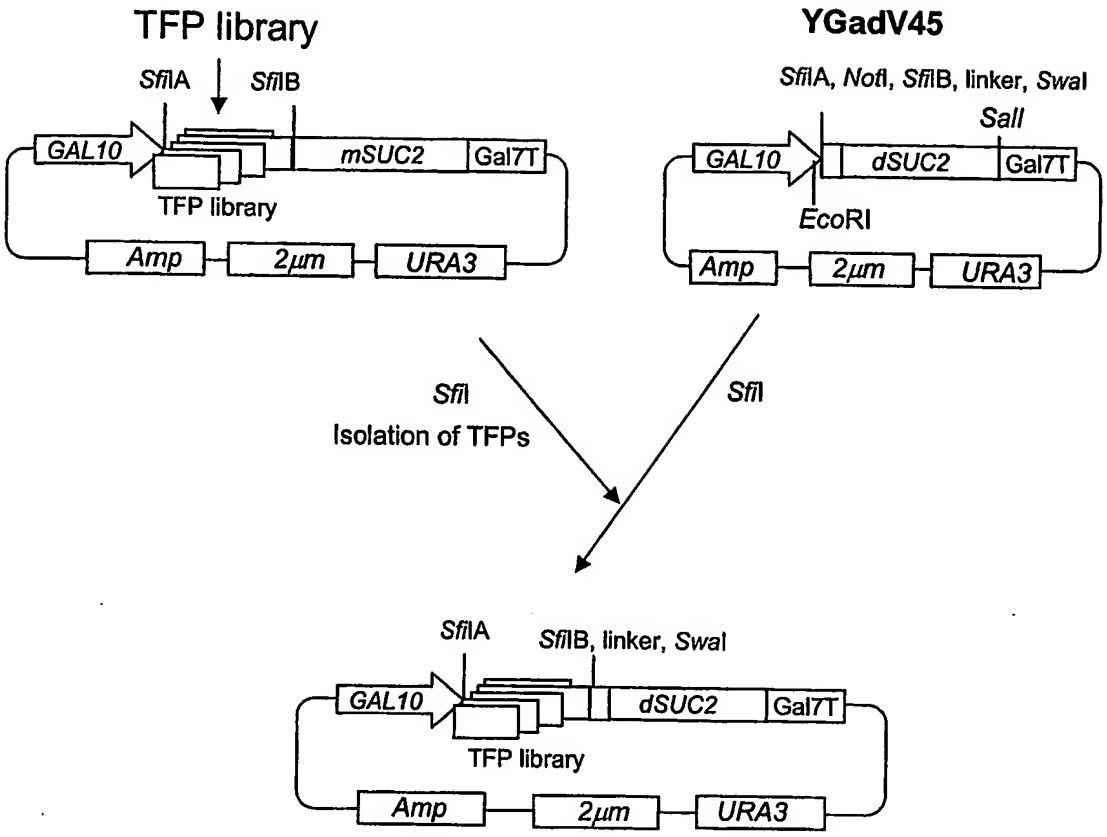

TFP library
SfiIA
SfiIB
GAL10
mSUC2
Gal7T
TFP library
Amp
2μm
URA3
YGadV45
SfiIA, NotI, SfiIB, linker, SwaI
SalI
GAL10
dSUC2
Gal7T
EcoRI
Amp
2μm
URA3
SfiI
Isolation of TFPs
SfiI
SfiIA
SfiIB, linker, SwaI
GAL10
dSUC2
Gal7T
TFP library
Amp
2μm
URA3

FIG. 11

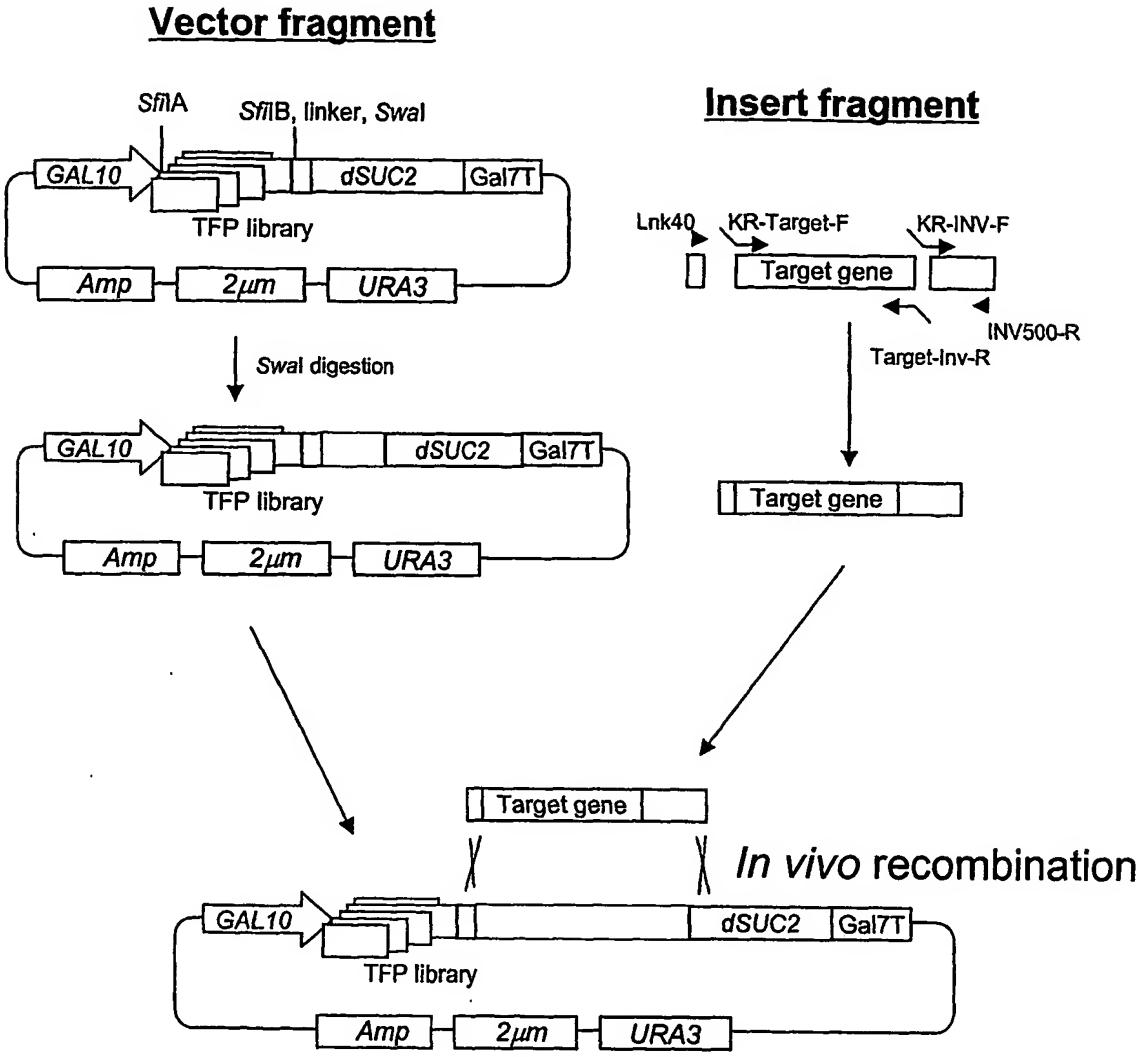

Vector fragment
SfiA
SfiB, linker, SwaI
GAL10
dSUC2
Gal7T
TFP library
Amp
2μm
URA3
SwaI digestion
Insert fragment
Lnk40
KR-Target-F
KR-INV-F
Target gene
INV500-R
Target-Inv-R
Target gene
Target gene
In vivo recombination
GAL10
dSUC2
Gal7T
TFP library
Amp
2μm
URA3

FIG. 12

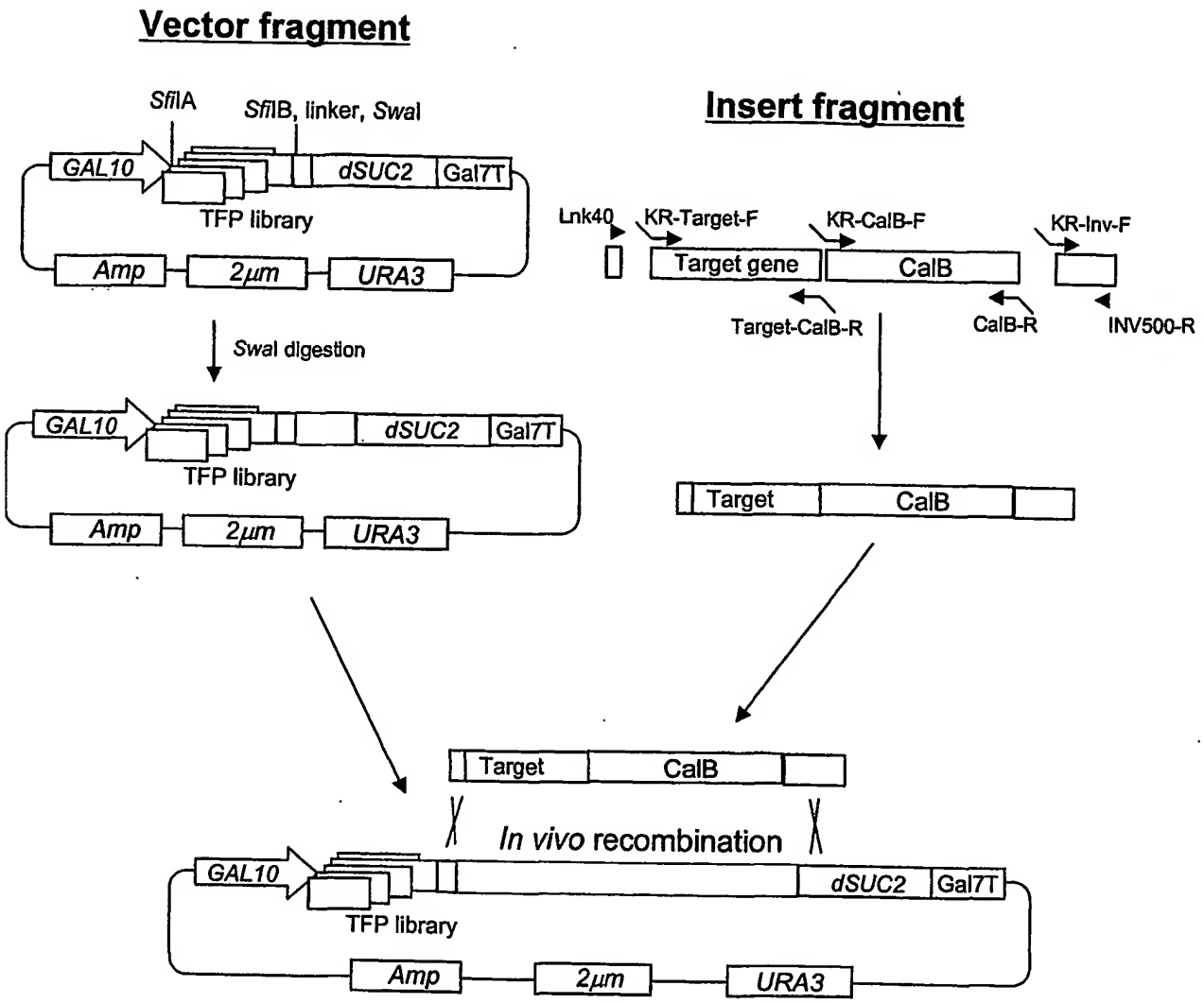

Vector fragment
SfiIA
SfiIB, linker, SwaI
GAL10
dSUC2
Gal7T
TFP library
Amp
2µm
URA3
SwaI digestion
Insert fragment
Lnk40
KR-Target-F
KR-CalB-F
KR-Inv-F
Target gene
CalB
Target-CalB-R
CalB-R
INV500-R
Target
CalB
In vivo recombination

FIG. 13

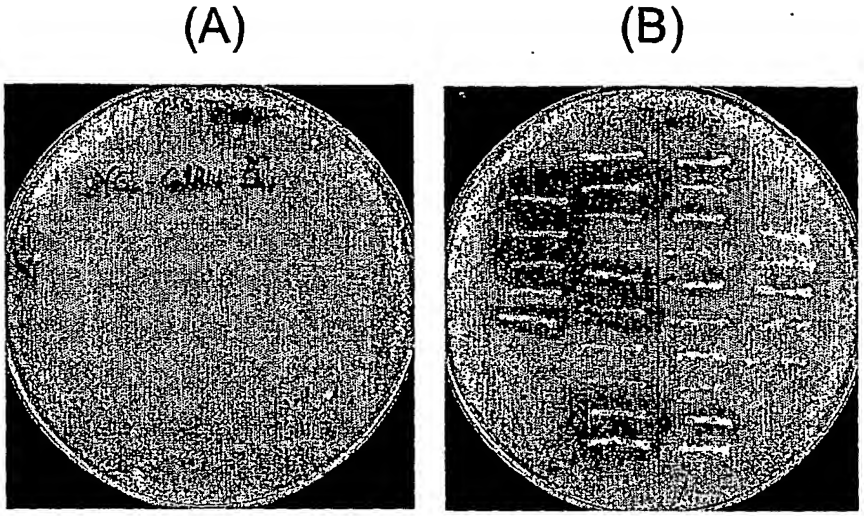


FIG. 14

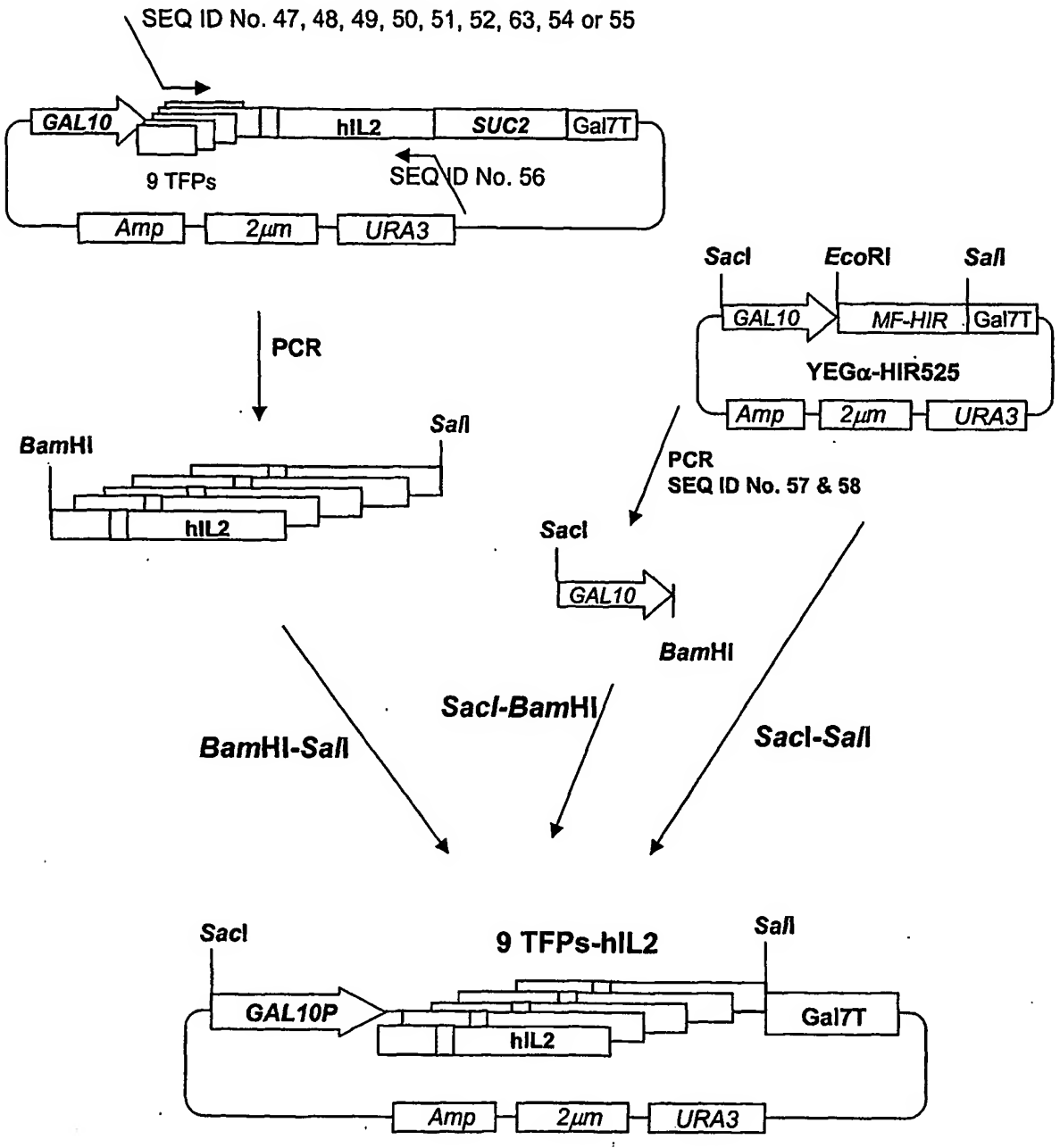

SEQ ID No. 47, 48, 49, 50, 51, 52, 63, 54 or 55
GAL10
hIL2
SUC2
Gal7T
9 TFPs
SEQ ID No. 56
Amp
2μm
URA3
SacI
EcoRI
SalI
GAL10
MF-HIR
Gal7T
YEGα-HIR525
Amp
2μm
URA3
PCR
BamHI
SalI
hIL2
PCR
SEQ ID No. 57 & 58
SacI
GAL10
BamHI
SacI-BamHI
BamHI-SalI
SacI-SalI
SacI
9 TFPs-hIL2
SalI
GAL10P
hIL2
Gal7T
Amp
2μm
URA3

FIG. 15

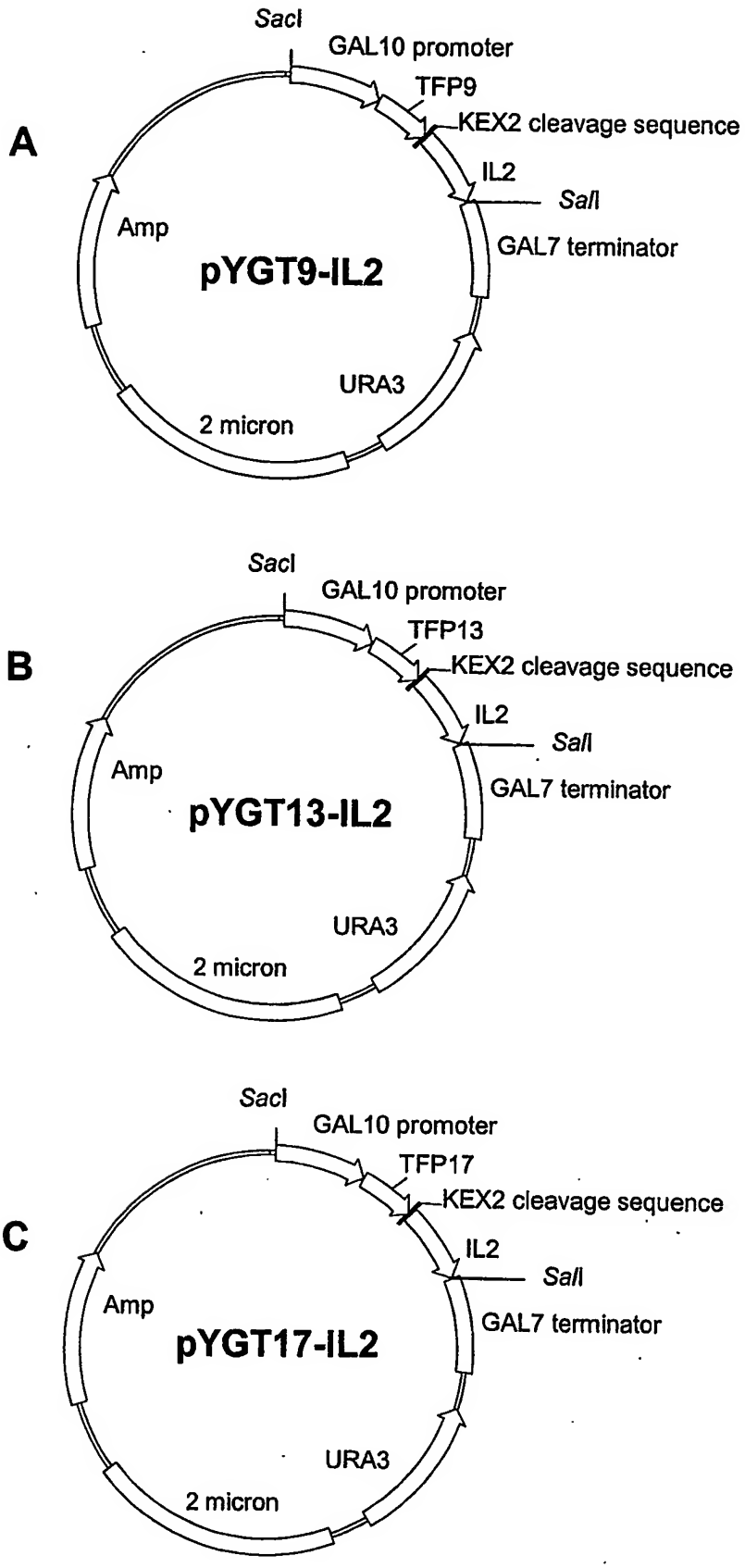

A
SacI
GAL10 promoter
TFP9
KEX2 cleavage sequence
IL2
SalI
GAL7 terminator
Amp
pYGT9-IL2
URA3
2 micron
B
SacI
GAL10 promoter
TFP13
KEX2 cleavage sequence
IL2
SalI
GAL7 terminator
Amp
pYGT13-IL2
URA3
2 micron
C
SacI
GAL10 promoter
TFP17
KEX2 cleavage sequence
IL2
SalI
GAL7 terminator
Amp
pYGT17-IL2
URA3
2 micron

FIG. 16

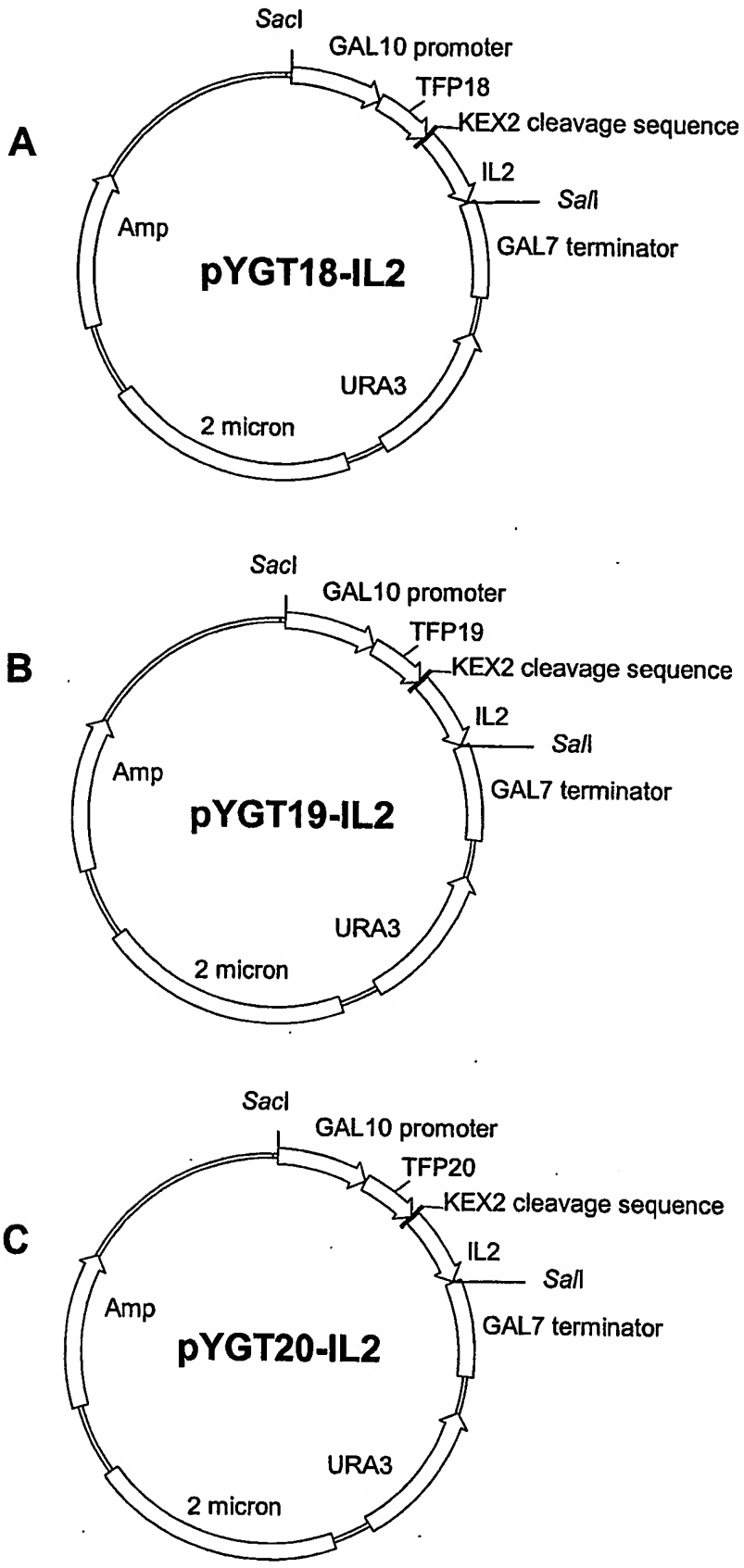

A
SacI
GAL10 promoter
TFP18
KEX2 cleavage sequence
IL2
SalI
GAL7 terminator
Amp
pYGT18-IL2
URA3
2 micron
B
SacI
GAL10 promoter
TFP19
KEX2 cleavage sequence
IL2
SalI
GAL7 terminator
Amp
pYGT19-IL2
URA3
2 micron
C
SacI
GAL10 promoter
TFP20
KEX2 cleavage sequence
IL2
SalI
GAL7 terminator
Amp
pYGT20-IL2
URA3
2 micron

FIG. 17

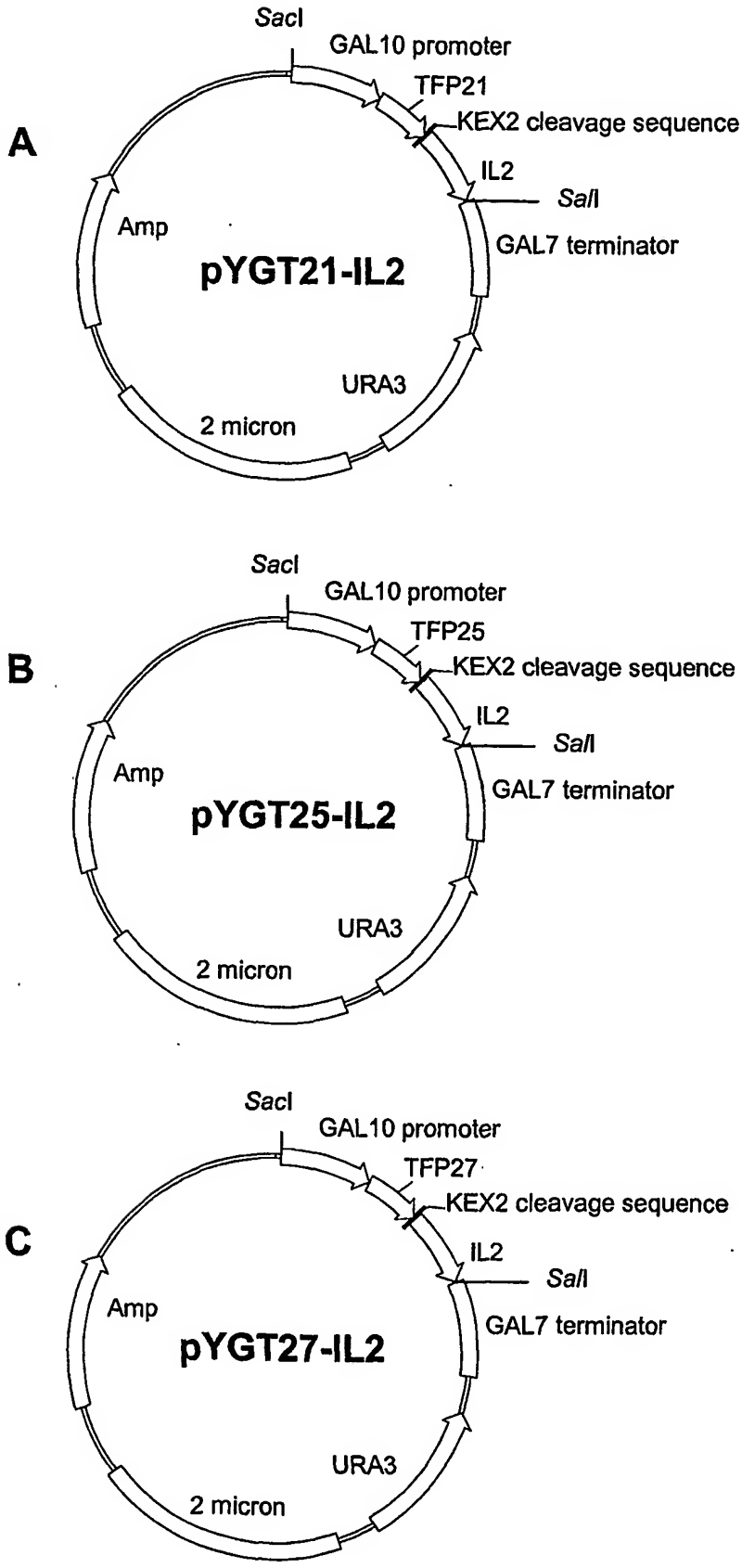

A
SacI
GAL10 promoter
TFP21
KEX2 cleavage sequence
IL2
SalI
GAL7 terminator
Amp
pYGT21-IL2
URA3
2 micron
B
SacI
GAL10 promoter
TFP25
KEX2 cleavage sequence
IL2
SalI
GAL7 terminator
Amp
pYGT25-IL2
URA3
2 micron
C
SacI
GAL10 promoter
TFP27
KEX2 cleavage sequence
IL2
SalI
GAL7 terminator
Amp
pYGT27-IL2
URA3
2 micron

FIG. 18

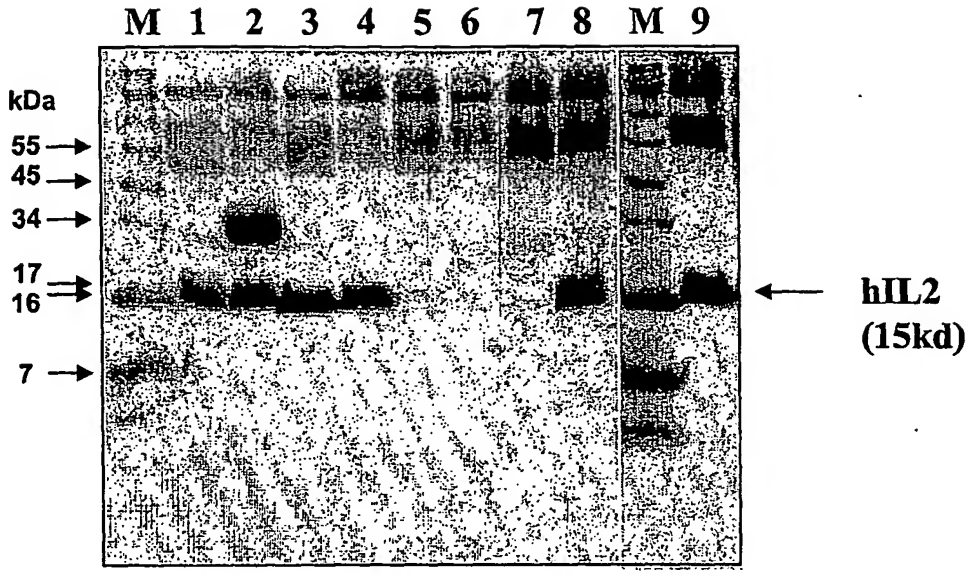


FIG. 19

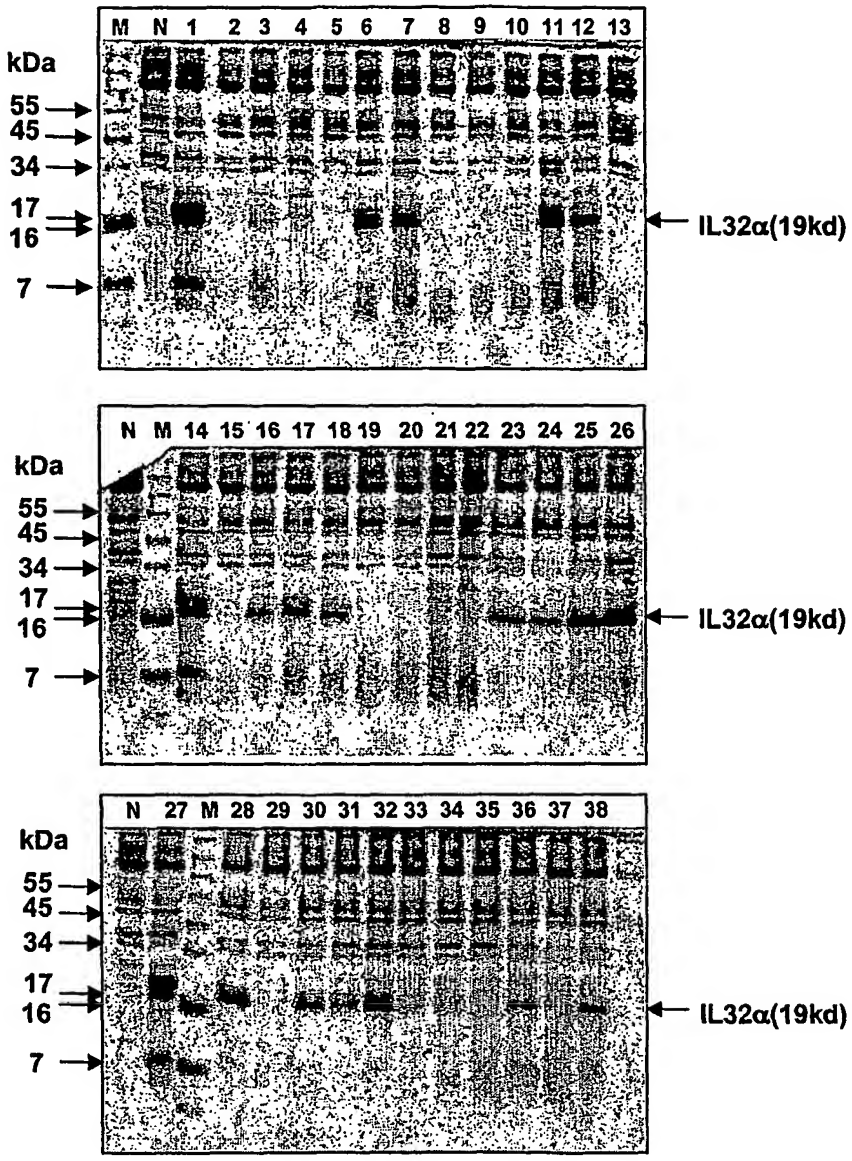

M N 1 2 3 4 5 6 7 8 9 10 11 12 13
kDa
55
45
34
17
16
7
IL32α(19kd)
N M 14 15 16 17 18 19 20 21 22 23 24 25 26
kDa
55
45
34
17
16
7
IL32α(19kd)
N 27 M 28 29 30 31 32 33 34 35 36 37 38
kDa
55
45
34
17
16
7
IL32α(19kd)

FIG.20

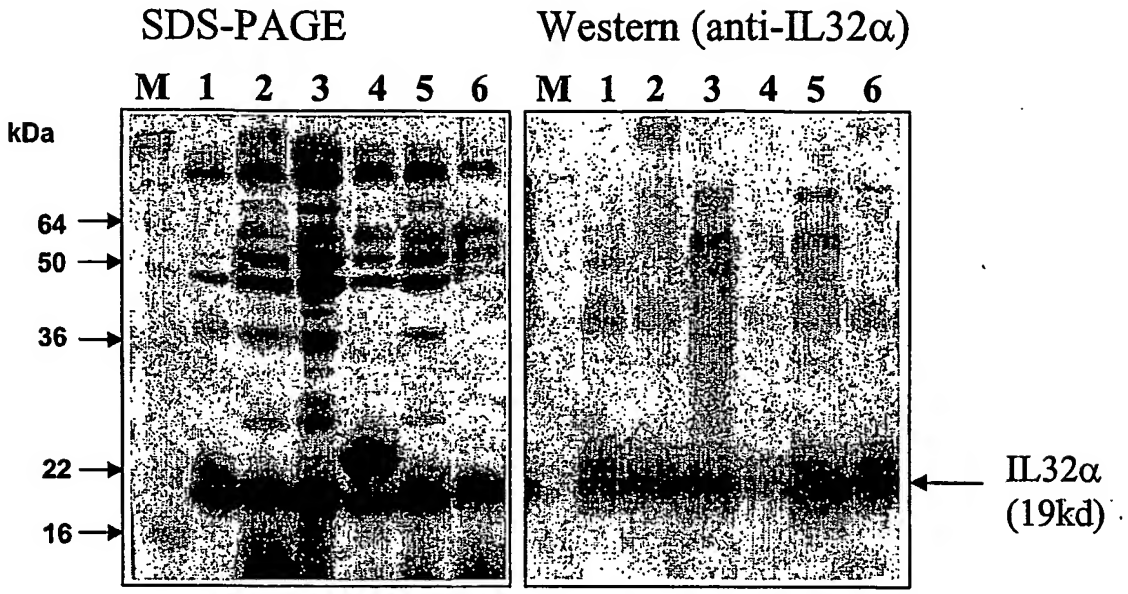
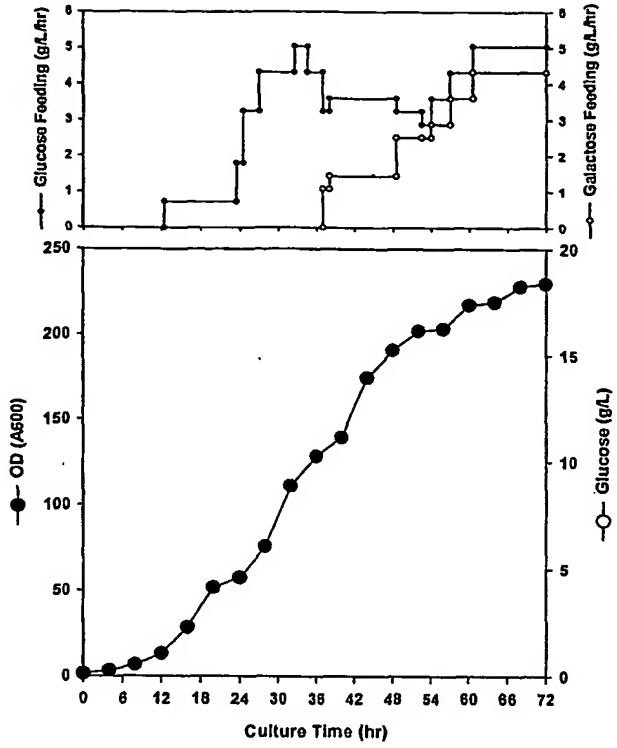

SDS-PAGE
Western (anti-IL32α)
M 1 2 3 4 5 6
M 1 2 3 4 5 6
kDa
64
50
36
22
16
IL32α
(19kd)

FIG.21

A


Glucose Feeding (g/L/hr)
Galactose Feeding (g/L/hr)
OD (A600)
Glucose (g/L)
Culture Time (hr)
0 6 12 18 24 30 36 42 48 54 60 66 72

B

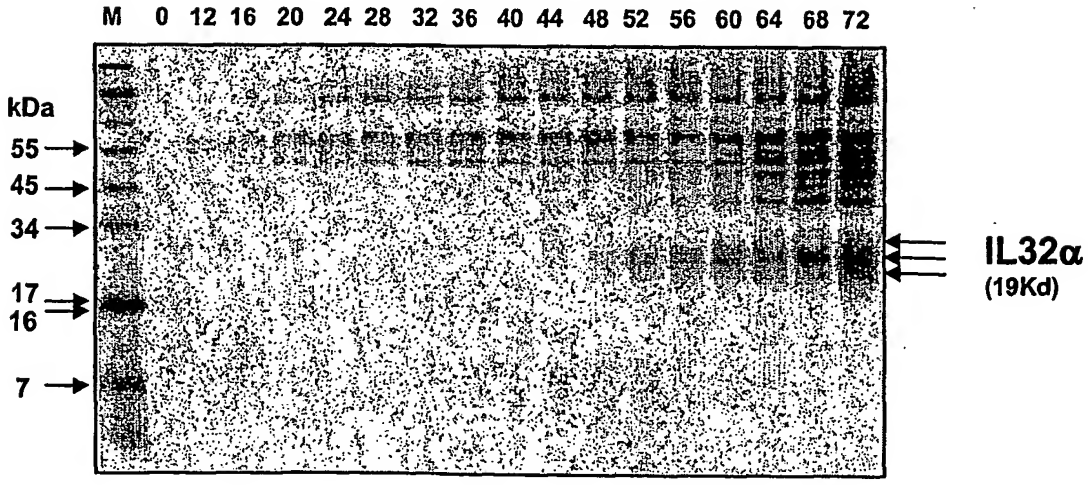

M 0 12 16 20 24 28 32 36 40 44 48 52 56 60 64 68 72
kDa
55
45
34
17
16
7
IL32α
(19Kd)

FIG.22

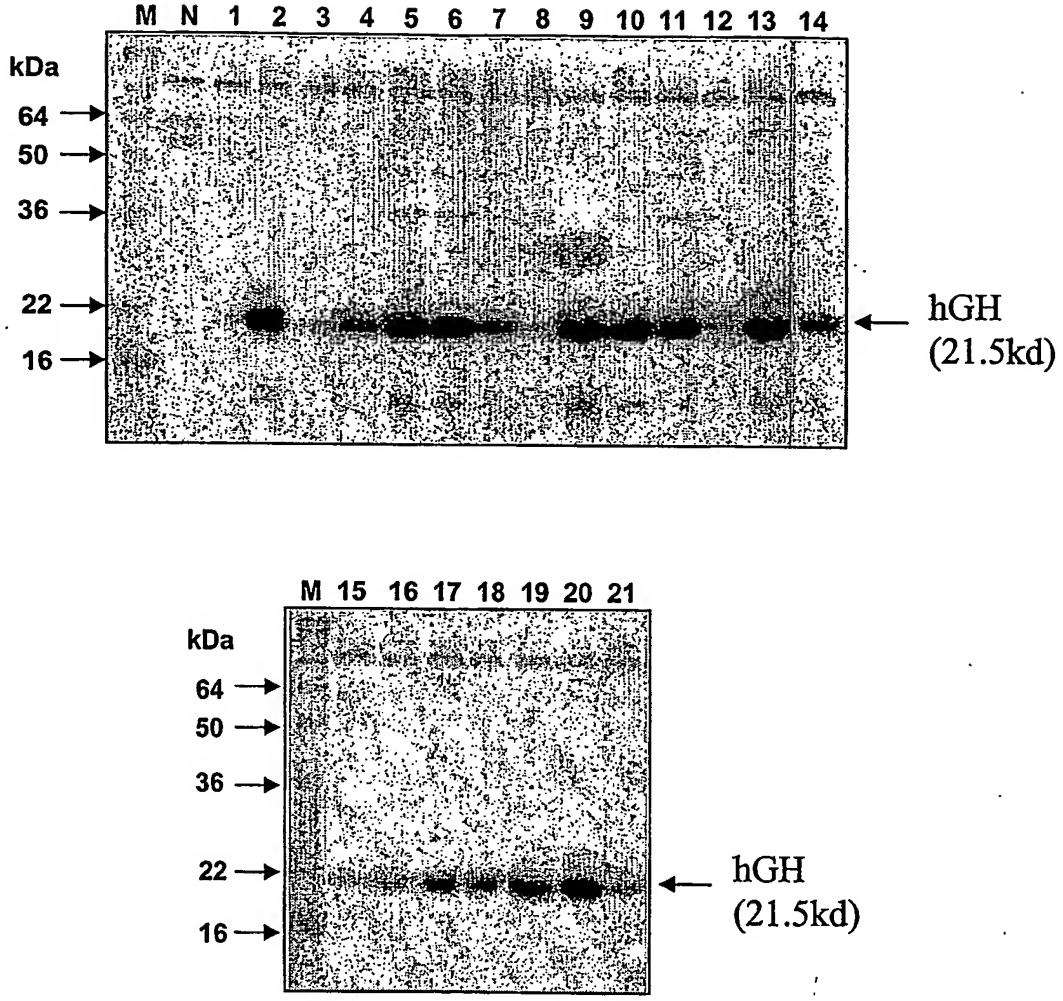
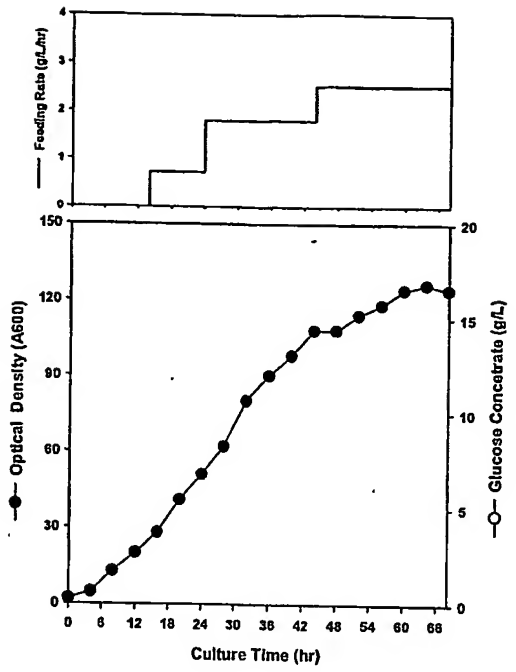

M N 1 2 3 4 5 6 7 8 9 10 11 12 13 14
kDa
64
50
36
22
16
hGH
(21.5kd)
M 15 16 17 18 19 20 21
kDa
64
50
36
22
16
hGH
(21.5kd)

FIG.23

A


Feeding Rate (g/L/hr)
Optical Density (A600)
Glucose Concentrate (g/L)
Culture Time (hr)
0 6 12 18 24 30 36 42 48 54 60 66
0 30 60 90 120 150
0 5 10 15 20
0 1 2 3 4

B

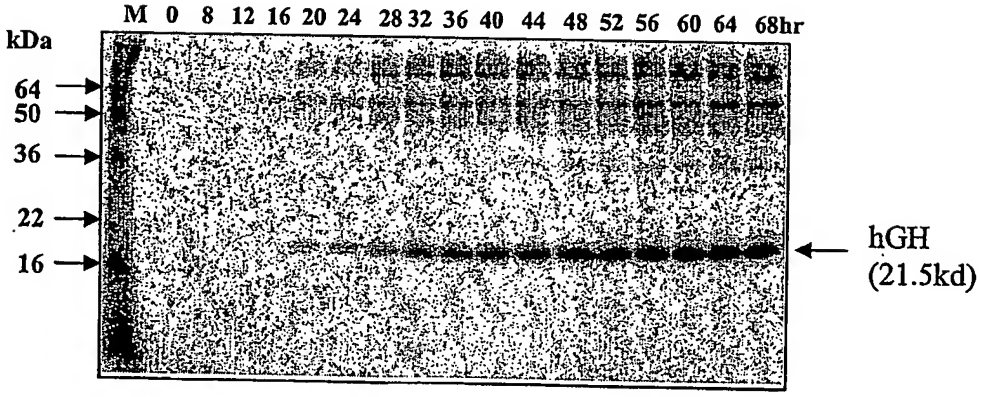

M 0 8 12 16 20 24 28 32 36 40 44 48 52 56 60 64 68hr
kDa
64
50
36
22
16
hGH
(21.5kd)

FIG.24

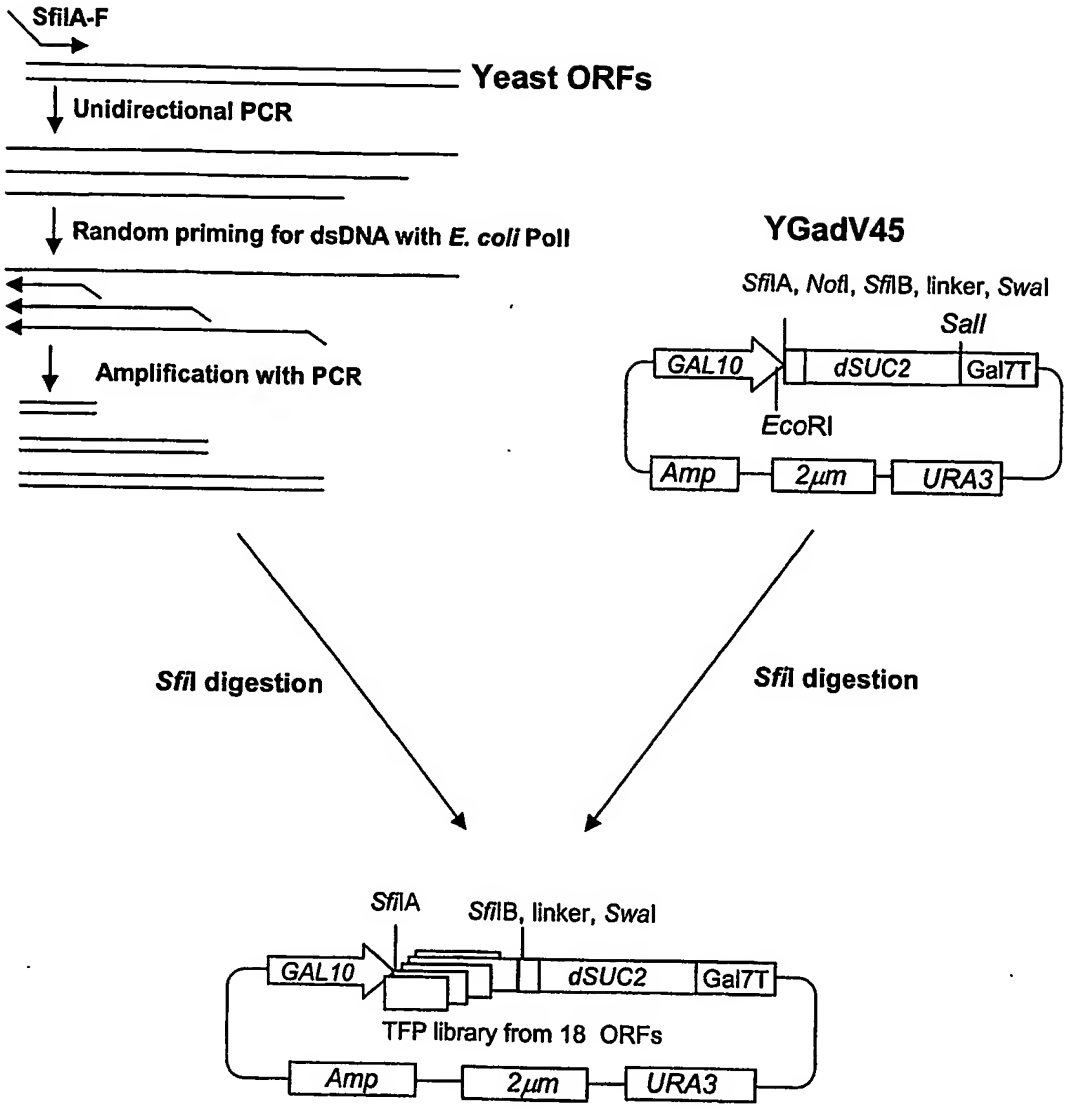

SfiIA-F
Yeast ORFs
Unidirectional PCR
Random priming for dsDNA with E. coli PolI
Amplification with PCR
YGadV45
SfiIA, NotI, SfiIB, linker, SwaI
SalI
GAL10
dSUC2
Gal7T
EcoRI
Amp
2μm
URA3
Sfil digestion
Sfil digestion
SfiIA
SfiIB, linker, SwaI
GAL10
dSUC2
Gal7T
TFP library from 18 ORFs
Amp
2μm
URA3

FIG.25

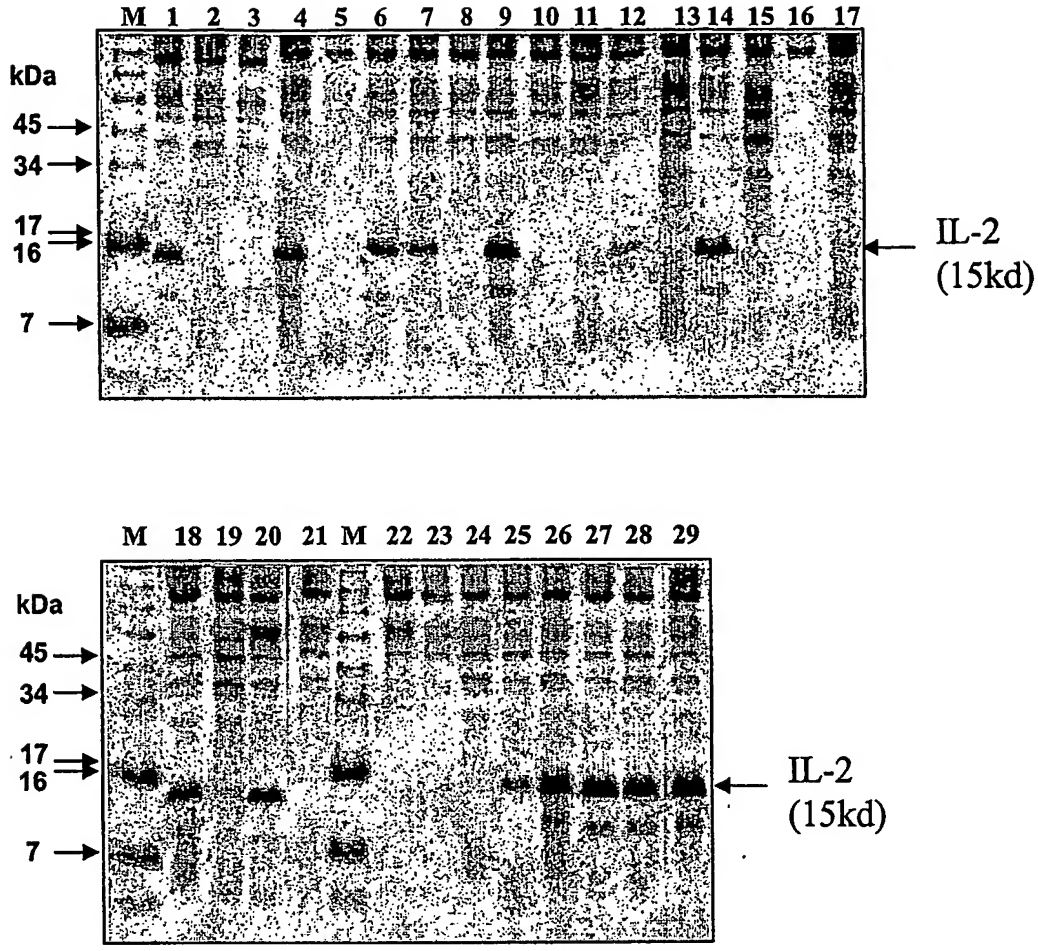

M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
kDa
45
34
17
16
7
IL-2
(15kd)
M 18 19 20 21 M 22 23 24 25 26 27 28 29
kDa
45
34
17
16
7
IL-2
(15kd)

FIG.26

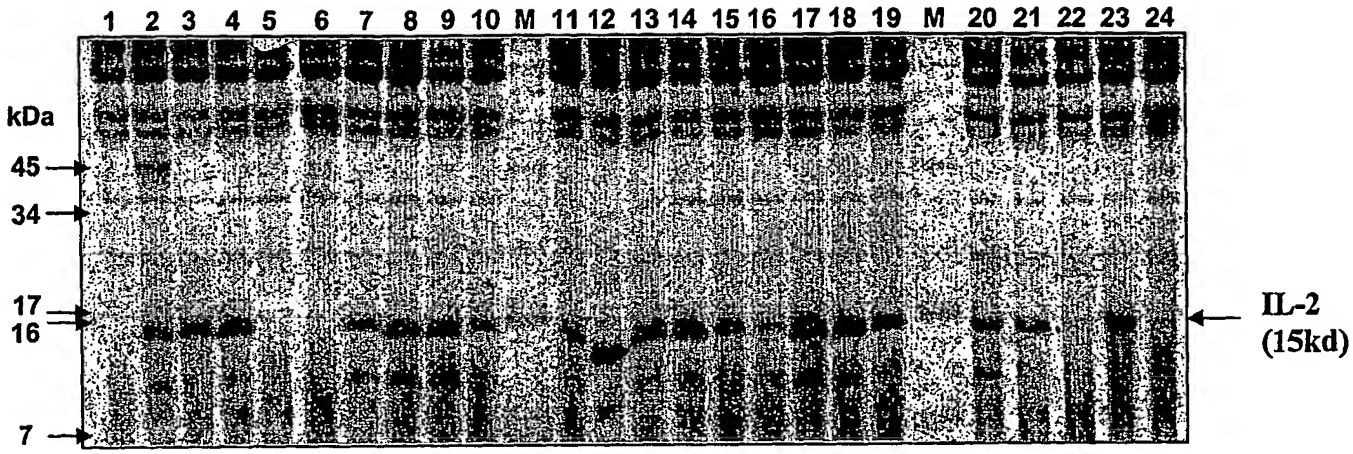

1 2 3 4 5 6 7 8 9 10 M 11 12 13 14 15 16 17 18 19 M 20 21 22 23 24
kDa
45
34
17
16
7
IL-2
(15kd)

FIG.27

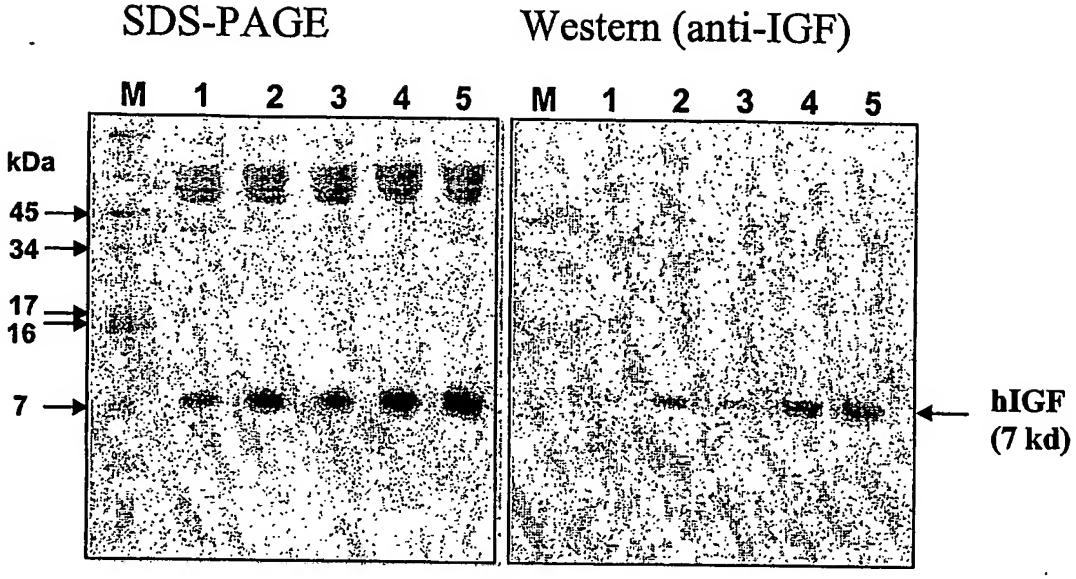

SDS-PAGE
Western (anti-IGF)
M 1 2 3 4 5 M 1 2 3 4 5
kDa
45
34
17
16
7
hIGF
(7 kd)

FIG.28

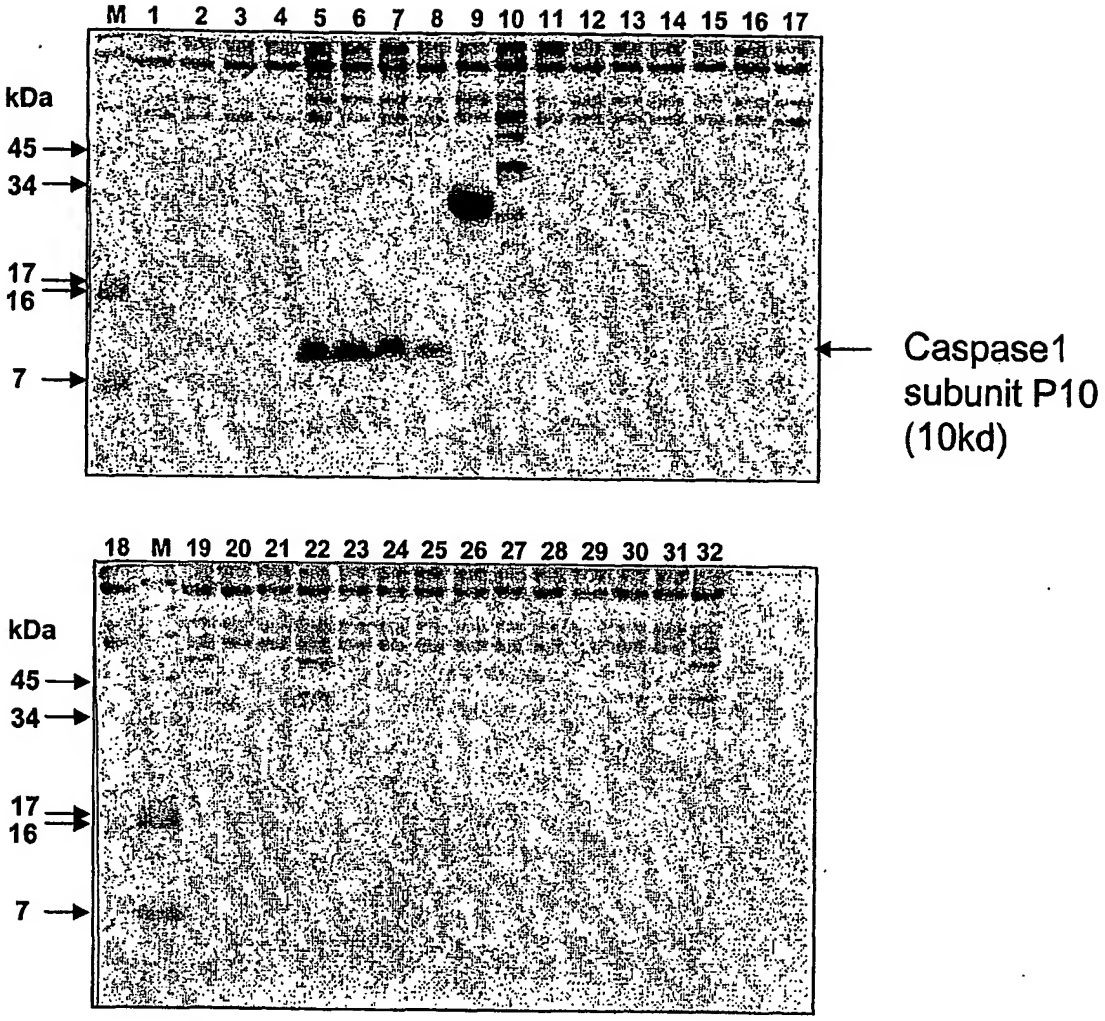

M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
kDa
45
34
17
16
7
Caspase1
subunit P10
(10kd)
18 M 19 20 21 22 23 24 25 26 27 28 29 30 31 32
kDa
45
34
17
16
7

FIG:29

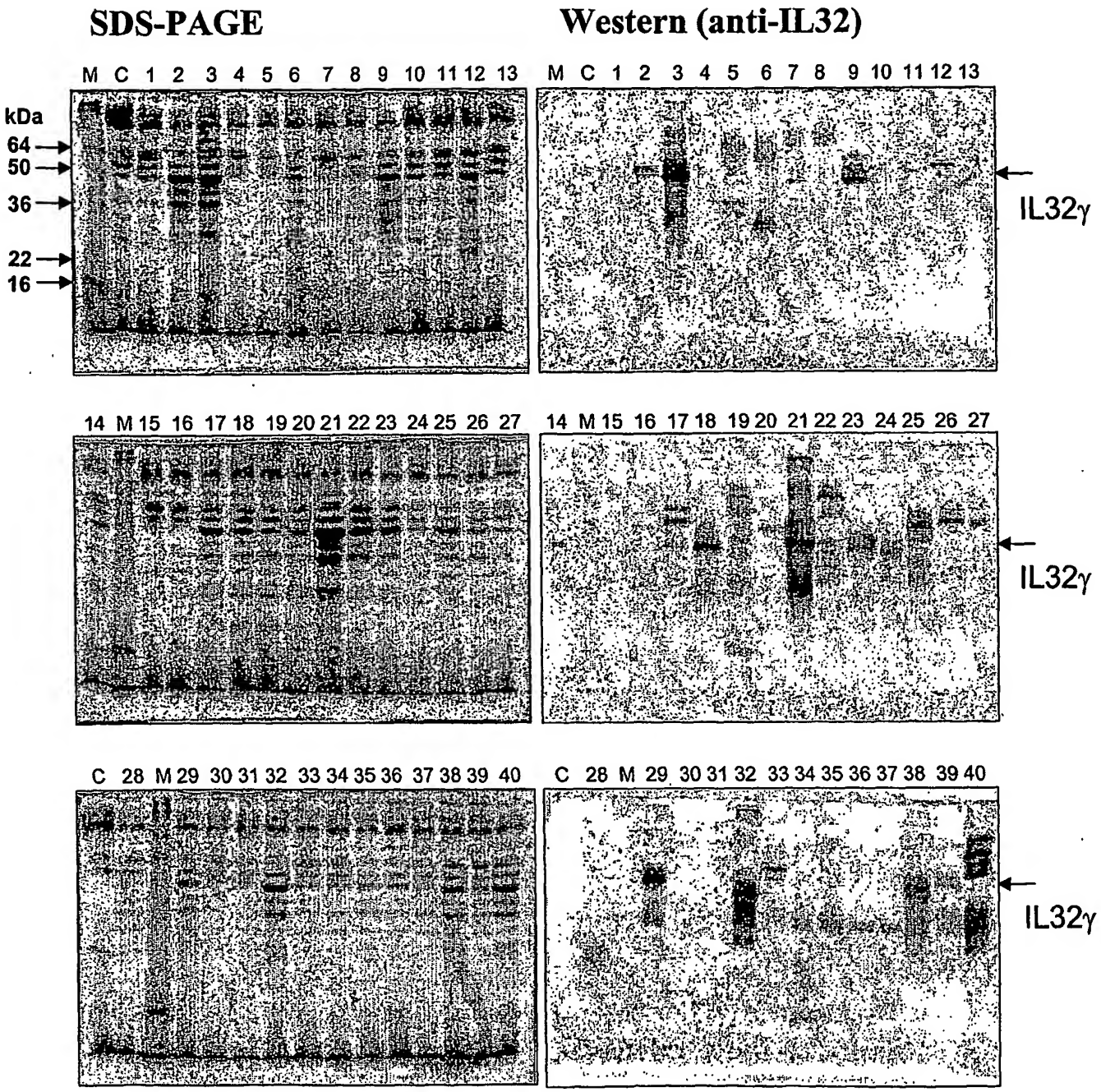
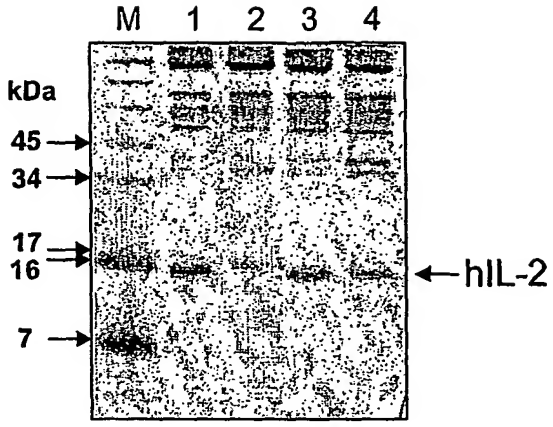

SDS-PAGE
Western (anti-IL32)
M C 1 2 3 4 5 6 7 8 9 10 11 12 13
kDa
64
50
36
22
16
IL32γ
14 M 15 16 17 18 19 20 21 22 23 24 25 26 27
IL32γ
C 28 M 29 30 31 32 33 34 35 36 37 38 39 40
IL32γ

FIG.30



SEQUENCE LISTING

<110> Korea Research Institute of Bioscience and Biotechnology
Sohn, Jung-Hoon
Choi, Eui-Sung
Bae, Jung-Hoon
Lee, Eung-Suck
Shin, Mi-Kyung
Yoon, Sung-Sook
Chun, Chang-Soo

<120> Library of Translational Fusion Partners for Producing Recombinant Proteins and Translational Fusion Partners Screened Therefrom

<130> 2472.002PC01/EKS/RAS

<160> 233

<170> PatentIn version 3.3

<210> 1
<211> 29
<212> DNA
<213> Artificial sequence

<220>
<223> SUC-F

<400> 1
gaattcaaaa atgcttttgc aagctttcc 29

<210> 2
<211> 30
<212> DNA
<213> Artificial sequence

<220>
<223> SUC-R

<400> 2
gtcgacttac tattttactt cccttacttg 30

<210> 3
<211> 24
<212> DNA
<213> Artificial sequence

<220>
<223> GAP-F

<400> 3
gagctcaagc ttaccagttc tcac 24

<210> 4
<211> 41
<212> DNA
<213> Artificial sequence

<220>
<223> SUCSS-R

<400> 4
gcggccgcac ggccgtaatg gcctgcagat attttggctg c 41

<210> 5
<211> 47
<212> DNA
<213> Artificial sequence

<220>
<223> SUM-F

<400> 5
gcggccgcct cggccctaga taaaaggtca atgacaaacg aaactag 47

<210> 6
<211> 33
<212> DNA
<213> Artificial sequence

<220>
<223> HSA-F

<400> 6
ggccattacg gccgtgatgc acacaagagt gag 33

<210> 7
<211> 33
<212> DNA
<213> Artificial sequence

<220>
<223> HSA-R

<400> 7
ggccgaggcg gcctaagcct aaggcagctt gac 33

<210> 8
<211> 35
<212> DNA
<213> Artificial sequence

<220>
<223> IL2-F

<400> 8
ggccattacg gccgtgcacc tacttcaagt tctac 35

<210> 9
<211> 33
<212> DNA
<213> Artificial sequence

<220>
<223> IL2-R

<400> 9
ggccgaggcg gccagttagt gttgagatga tgc 33

<210> 10
<211> 58
<212> DNA
<213> Artificial sequence

<220>
<223> SfiI-SUC-F

<400> 10
gaattcaaaa ggccattacg gccgcggccg cctcggccct agataaaagg tcaatgac 58

<210> 11
<211> 29
<212> DNA
<213> Artificial sequence

<220>
<223> SUC-Xho-R

<400> 11
ggctcgagct attttacttc ccttacttg 29

<210> 12
<211> 26
<212> DNA
<213> Artificial sequence

<220>
<223> GAL100-F

<400> 12
gatatgtata tggtggtaat gccatg 26

<210> 13
<211> 41
<212> DNA
<213> Artificial sequence

<220>
<223> Xho-F0-R

<400> 13
ctagggccga ggcggccctc gagggccgta atggcctttt g 41

<210> 14
<211> 42
<212> DNA
<213> Artificial sequence

<220>

<223> Xho-F1-R

<400> 14
ctagggccga ggcggccgct cgagggccgt aatggccttt tg 42

<210> 15
<211> 43
<212> DNA
<213> Artificial sequence

<220>
<223> Xho-F2-R

<400> 15
ctagggccga ggcggccgtc tcgagggccg taatggcctt ttg 43

<210> 16
<211> 30
<212> DNA
<213> Artificial sequence

<220>
<223> ASA24N6

<220>
<221> misc_feature
<222> (25)..(30)
<223> n is a, c, g, or t

<400> 16
gccagcagag gccgaggcgg ccagnnnnnn 30

<210> 17
<211> 24
<212> DNA
<213> Artificial sequence

<220>
<223> ASA24

<400> 17
gccagcagag gccgaggcgg ccag 24

<210> 18
<211> 65
<212> DNA
<213> Artificial sequence

<220>
<223> INV45-F

<400> 18
gcggccgcct cggcctctgc tggcctcgcc ttagataaaa gatttaaatg acaccgtatg 60

gggta 65

<210> 19
<211> 35
<212> DNA
<213> Artificial sequence

<220>
<223> KR-target-F

<220>
<221> misc_feature
<222> (19)..(35)
<223> n is a, c, g, or t

<400> 19
ctcgccttag ataaaagann nnnnnnnnnn nnnnn 35

<210> 20
<211> 35
<212> DNA
<213> Artificial sequence

<220>
<223> Target-INV-R

<220>
<221> misc_feature
<222> (19)..(35)
<223> n is a, c, g, or t

<400> 20
cattgaacgc ttgtccaann nnnnnnnnnn nnnnn 35

<210> 21
<211> 38
<212> DNA
<213> Artificial sequence

<220>
<223> KR-Inv-F

<400> 21
ttggacaagc gttcaatgac aaacgaaact agcgatag 38

<210> 22
<211> 25
<212> DNA
<213> Artificial sequence

<220>
<223> Inv500-R

<400> 22
tcataatcca tttttgagaa ggttc 25

<210> 23
<211> 40
<212> DNA
<213> Artificial sequence

<220>
<223> LNK40

<400> 23
ggccgcctcg gcctctgctg gcctcgcctt agataaaaga 40

<210> 24
<211> 30
<212> DNA
<213> Artificial sequence

<220>
<223> KR-CalB-F

<400> 24
ttggacaagc gtctaccttc cggttcggac 30

<210> 25
<211> 38
<212> DNA
<213> Artificial sequence

<220>
<223> CalB-Inv-R

<400> 25
cattgaacgc ttgtccaagg gggtgacgat gccggagc 38

<210> 26
<211> 35
<212> DNA
<213> Artificial sequence

<220>
<223> Target-CalB-R

<220>
<221> misc_feature
<222> (19)..(35)
<223> n is a, c, g, or t

<400> 26
aggtagacgc ttgtccaann nnnnnnnnnn nnnnn 35

<210> 27
<211> 38
<212> DNA
<213> Artificial sequence

<220>
<223> KR-IL2-F

<400> 27
ctcgccttag ataaaagagc acctacttca agttctac 38

<210> 28
<211> 38
<212> DNA
<213> Artificial sequence

<220>
<223> IL2-INV-R

<400> 28
cattgaacgc ttgtccaaag ttagtgttga gatgatgc 38

<210> 29
<211> 217
<212> PRT
<213> Artificial sequence

<220>
<223> TFP9-AA

<400> 29

Met Val Phe Gly Gln Leu Tyr Ala Leu Phe Ile Phe Thr Leu Ser Cys
1 5 10 15

Cys Ile Ser Lys Thr Val Gln Ala Asp Ser Ser Lys Glu Ser Ser Ser
20 25 30

Phe Ile Ser Phe Asp Lys Glu Ser Asn Trp Asp Thr Ile Ser Thr Ile
35 40 45

Ser Ser Thr Ala Asp Val Ile Ser Ser Val Asp Ser Ala Ile Ala Val
50 55 60

Phe Glu Phe Asp Asn Phe Ser Leu Leu Asp Asn Leu Met Ile Asp Glu
65 70 75 80

Glu Tyr Pro Phe Phe Asn Arg Phe Phe Ala Asn Asp Val Ser Leu Thr
85 90 95

Val His Asp Asp Ser Pro Leu Asn Ile Ser Gln Ser Leu Ser Pro Ile
100 105 110

Met Glu Gln Phe Thr Val Asp Glu Leu Pro Glu Ser Ala Ser Asp Leu
115 120 125

Leu Tyr Glu Tyr Ser Leu Asp Asp Lys Ser Ile Val Leu Phe Lys Phe
130 135 140

Thr Ser Asp Ala Tyr Asp Leu Lys Lys Leu Asp Glu Phe Ile Asp Ser
145 150 155 160

Cys Leu Ser Phe Leu Glu Asp Lys Ser Gly Asp Asn Leu Thr Val Val
165 170 175

Ile Asn Ser Leu Gly Trp Ala Phe Glu Asp Glu Asp Gly Asp Asp Glu
180 185 190

Tyr Ala Thr Glu Glu Thr Leu Ser His His Asp Asn Asn Lys Gly Lys
195 200 205

Glu Gly Asp Asp Leu Ala Ala Ser Ala
210 215

<210> 30
<211> 728
<212> DNA
<213> Artificial sequence

<220>
<223> TFP9-nt

<400> 30
ggccattacg gccggggatt gataataacc actgctgtga ctatatataa taagaatcga 60

actgtaaagt taaagcaatg gtgttcggtc agctgtatgc ccttttcatc ttcacgttat 120

catgttgtat ttccaaaact gtgcaagcag attcatccaa ggaaagctct tcctttattt 180

cgttcgacaa agagagtaac tgggatacca tcagcactat atcttcaacg gcagatgtta 240

tatcatccgt tgacagtgct atcgctgttt ttgaatttga caatttctca ttattggaca 300

acttgatgat tgacgaagaa tacccattct tcaatagatt ctttgccaat gatgtcagtt 360

taactgttca tgacgattcg cctttgaaca tctctcaatc attatctccc attatggaac 420

aatttactgt ggatgaatta cctgaaagtg cctctgactt actatatgaa tactccttag 480

atgataaaag catcgttttg ttcaagttta cctcggatgc ctacgatttg aaaaaattag 540

atgaatttat tgattcttgc ttatcgtttt tggaggataa atctggcgac aatttgactg 600

tggttattaa ctctcttggt tgggcttttg aagatgaaga tggtgacgat gaatatgcaa 660

cagaagagac tttgagccat catgataaca acaagggtaa agaaggcgac gatctggccg 720

cctcggcc 728

<210> 31
<211> 127
<212> PRT

<213> Artificial sequence

<220>
<223> TFP13-aa

<400> 31

Met Lys Phe Ser Thr Ala Val Thr Thr Leu Ile Ser Ser Gly Ala Ile
1 5 10 15

Val Ser Ala Leu Pro His Val Asp Val His Gln Glu Asp Ala His Gln
20 25 30

His Lys Arg Ala Val Ala Tyr Lys Tyr Val Tyr Glu Thr Val Val Val
35 40 45

Asp Ser Asp Gly His Thr Val Thr Pro Ala Ala Ser Glu Val Ala Thr
50 55 60

Ala Ala Thr Ser Ala Ile Ile Thr Thr Ser Val Leu Ala Pro Thr Ser
65 70 75 80

Ser Ala Ala Ala Gly Ile Ala Ala Ser Ile Ala Val Ser Ser Ala Ala
85 90 95

Leu Ala Lys Asn Glu Lys Ile Ser Asp Ala Ala Ala Ser Ala Thr Ala
100 105 110

Ser Thr Ser Gln Gly Ala Ser Ser Ser Ser Leu Ala Ala Ser Ala
115 120 125

<210> 32
<211> 618
<212> DNA
<213> Artificial sequence

<220>
<223> TFP13-aa

<220>
<221> misc_feature
<222> (42)..(42)
<223> n is a, c, g, or t

<400> 32
ggccattacg gccggggatt caaatatata tatctactca gnttgaataa gacactatag 60

caagaccatt tgaactgaaa gaaacagttt ctttgctccc ctctcgaatt ccaactattt 120

acagtccttc ctttataaaa attaactagc gagcaagaaa acatttgttt agtgctaccc 180

aactacttac attcctttaa aaaccacaat atttaagtta acctgagctt tatttttaaa 240

atgaaattct caactgccgt tactacgttg attagttctg gtgccatcgt gtctgcttta 300

ccacacgtgg atgttcacca agaagatgcc caccaacata agagggccgt tgcgtacaaa 360

tacgtttacg aaactgttgt tgtcgattct gatggccaca ctgtaactcc tgctgcttca 420

gaagtcgcta ctgctgctac ctctgctatc attacaacat ctgtgttggc tccaacctcc 480

tccgcagccg ctgggatagc cgcttccatt gctgtttcat ctgctgcctt agccaagaat 540

gagaaaatct ctgatgccgc tgcatctgcc actgcctcaa catctcaagg ggcatcctcc 600

tcctccctgg ccgcctcg 618

<210> 33
<211> 68
<212> PRT
<213> Artificial sequence

<220>
<223> TFP17-AA

<400> 33

Met Lys Phe Ser Ser Val Thr Ala Ile Thr Leu Ala Thr Val Ala Thr
1 5 10 15

Val Ala Thr Ala Lys Lys Gly Glu His Asp Phe Thr Thr Thr Leu Thr
20 25 30

Leu Ser Ser Asp Gly Ser Leu Thr Thr Thr Thr Ser Thr His Thr Thr
35 40 45

His Lys Tyr Gly Lys Phe Asn Lys Thr Ser Lys Ser Lys Thr Pro Trp
50 55 60

Ala Ala Ser Ala
65

<210> 34
<211> 391
<212> DNA
<213> Artificial sequence

<220>
<223> TFP17-nt

<400> 34
ggccattacg gccggggttt ctttctcttt tttcttttt gaataaagaa ttttccttta 60

aggagtaact taagcattta gctgcacatt aaacactttt ttttttactt ctaactcaca 120

cacttttgga agaacattta ttttttcgac cttctttccc aaatacccag cgctttataa 180

ttgaaatatg aagttctctt ctgttactgc tattactcta gccaccgttg ccaccgttgc 240

cactgctaag aagggtgaac atgatttcac taccacttta actttgtcat cggacggtag 300

tttaactact accacctcta ctcataccac tcacaagtat ggtaagttca acaagacttc 360

caagtccaag acccccctggg ccgcctcggc c 391

<210> 35
<211> 199
<212> PRT
<213> Artificial sequence

<220>
<223> TFP18-aa

<400> 35

Met Gln Phe Lys Asn Ala Leu Thr Ala Thr Ala Ile Leu Ser Ala Ser
1 5 10 15

Ala Leu Ala Asn Ser Thr Thr Ser Ile Pro Ser Ser Cys Ser Ile Gly
20 25 30

Thr Ser Ala Thr Ala Thr Ala Gln Ala Asp Leu Asp Lys Ile Ser Gly
35 40 45

Cys Ser Thr Ile Val Gly Asn Leu Thr Ile Thr Gly Asp Leu Gly Ser
50 55 60

Ala Ala Leu Ala Ser Ile Gln Glu Ile Asp Gly Ser Leu Thr Ile Phe
65 70 75 80

Asn Ser Ser Ser Leu Ser Ser Phe Ser Ala Asp Ser Ile Lys Lys Ile
85 90 95

Thr Gly Asp Leu Asn Met Gln Glu Leu Ile Ile Leu Thr Ser Ala Ser
100 105 110

Phe Gly Ser Leu Gln Glu Val Asp Ser Ile Asn Met Val Thr Leu Pro
115 120 125

Ala Ile Ser Thr Phe Ser Thr Asp Leu Gln Asn Ala Asn Asn Ile Ile
130 135 140

Val Ser Asp Thr Thr Leu Glu Ser Val Glu Gly Phe Ser Thr Leu Lys
145 150 155 160

Lys Val Asn Val Phe Asn Ile Asn Asn Asn Arg Tyr Leu Asn Ser Phe
165 170 175

Gln Ser Ser Leu Glu Ser Val Ser Asp Ser Leu Gln Phe Ser Ser Asn
180 185 190

Gly Asp Leu Ala Ala Ser Ala
195

<210> 36
<211> 760
<212> DNA
<213> Artificial sequence

<220>
<223> TFP18-nt

<400> 36
ggccattacg gccgaaaaga aattaaataa aagagttaat cgttcattcg cttctacaca 60

gtttaatctt ttccattttt ctttcaacaa gtccctttga gctatcaaga atacgtttat 120

ttgactttta aagatctagt tttaatttta ctattattcc gcaatgcaat tcaagaacgc 180

tttgactgct actgctattc taagtgcctc cgctctagct aactcaacta cttctattcc 240

atcttcatgt agtattggta cttctgccac tgctactgct caagctgatt tggacaaaat 300

ctccggttgt agtaccattg ttggtaactt gaccatcacc ggtgacttgg gttccgctgc 360

tttggctagt atccaagaga ttgatggttc cttgactatc ttcaactcca gttctttatc 420

ttctttctcc gctgactcta tcaagaaaat caccggtgat ttgaacatgc aagaattgat 480

cattttgacc agtgcttctt tcggttcttt gcaagaagta gactccatta acatggtgac 540

tttgcctgcc atttctacct tctccaccga tttacaaaat gctaacaaca ttattgtttc 600

tgacaccact ttggaaagtg tcgaaggttt ctccactttg aagaaggtta atgtttttaa 660

catcaacaac aacagatatc taaactcttt ccaatcttcc ttggaaagtg tctctgactc 720

tttacaattc tcttccaacg gtgacctggc cgcctcggcc 760

<210> 37
<211> 148
<212> PRT
<213> Artificial sequence

<220>
<223> TFP19-aa

<400> 37

Met Val Ser Lys Thr Trp Ile Cys Gly Phe Ile Ser Ile Ile Thr Val
1 5 10 15

Val Gln Ala Leu Ser Cys Glu Lys His Asp Val Leu Lys Lys Tyr Gln

20 25 30

Val Gly Lys Phe Ser Ser Leu Thr Ser Thr Glu Arg Asp Thr Pro Pro
35 40 45

Ser Thr Thr Ile Glu Lys Trp Trp Ile Asn Val Cys Glu Glu His Asn
50 55 60

Val Glu Pro Pro Glu Glu Cys Lys Lys Asn Asp Met Leu Cys Gly Leu
65 70 75 80

Thr Asp Val Ile Leu Pro Gly Lys Asp Ala Ile Thr Thr Gln Ile Ile
85 90 95

Asp Phe Asp Lys Asn Ile Gly Phe Asn Val Glu Glu Thr Glu Ser Ala
100 105 110

Leu Thr Leu Thr Leu Lys Gly Ala Thr Trp Gly Ala Asn Ser Phe Asp
115 120 125

Ala Lys Leu Glu Phe Gln Cys Asn Asp Asn Met Lys Gln Asp Glu Leu
130 135 140

Ala Ala Ser Ala
145

<210> 38
<211> 464
<212> DNA
<213> Artificial sequence

<220>
<223> TFP19-nt

<400> 38
ggccattacg gccggggacg atggtatcga agacttggat atgtggcttc atcagtataa 60

ttacagtggt acaggccttg tcctgcgaga agcatgatgt attgaaaaag tatcaggtgg 120

gaaaatttag ctcactaact tctacggaaa gggatactcc gccaagcaca actattgaaa 180

agtggtggat aaacgtttgc gaagagcata acgtagaacc tcctgaagaa tgtaaaaaaa 240

atgacatgct atgtggttta acagatgtca tcttgcccgg taaggatgct atcaccactc 300

aaattataga ttttgacaaa aacattggct tcaatgtcga ggaaactgag agtgcgctta 360

cattgacact aaaaggcgct acgtggggcg ccaattcttt tgacgcaaaa ctagaatttc 420

agtgtaatga caatatgaaa caagacgaac tggccgcctc ggcc 464

<210> 39
<211> 187
<212> PRT
<213> Artificial sequence

<220>
<223> TFP20-aa

<400> 39

Met Leu Phe Lys Ser Leu Ser Lys Leu Ala Thr Ala Ala Ala Phe Phe
1 5 10 15

Ala Gly Val Ala Thr Ala Asp Asp Val Pro Ala Ile Glu Val Val Gly
20 25 30

Asn Lys Phe Phe Tyr Ser Asn Asn Gly Ser Gln Phe Tyr Ile Arg Gly
35 40 45

Val Ala Tyr Gln Ala Asp Thr Ala Asn Glu Thr Ser Gly Ser Thr Val
50 55 60

Asn Asp Pro Leu Ala Asn Tyr Glu Ser Cys Ser Arg Asp Ile Pro Tyr
65 70 75 80

Leu Lys Lys Leu Asn Thr Asn Val Ile Arg Val Tyr Ala Ile Asn Thr
85 90 95

Thr Leu Asp His Ser Glu Cys Met Lys Ala Leu Asn Asp Ala Asp Ile
100 105 110

Tyr Val Ile Ala Asp Leu Ala Ala Pro Ala Thr Ser Ile Asn Arg Asp
115 120 125

Asp Pro Thr Trp Thr Val Asp Leu Phe Asn Ser Tyr Lys Thr Val Val
130 135 140

Asp Thr Phe Ala Asn Tyr Thr Asn Val Leu Gly Phe Phe Ala Gly Asn
145 150 155 160

Glu Val Thr Asn Asn Tyr Thr Asn Thr Asp Ala Ser Ala Phe Val Lys
165 170 175

Ala Ala Ile Arg Asp Val Leu Ala Ala Ser Ala
180 185

<210> 40
<211> 664
<212> DNA

<213> Artificial sequence

<220>
<223> TFP20-nt

<400> 40
ggccattacg gccggggtgt cgttttatta agctatttca aaatcagttt ttatttttaa 60

agtctgataa aacaaaaaca acaaacacag ctaaatctca acaatgttgt ttaaatccct 120

ttcaaagtta gcaaccgctg ctgctttttt tgctggcgtc gcaactgcgg acgatgttcc 180

agcgattgaa gttgttggta ataagttttt ctactccaac aacggtagtc agttctacat 240

aagaggtgtt gcttatcagg ctgataccgc taatgaaact agcggatcta ctgtcaacga 300

tcctttggcc aattatgaga gttgttccag agatattcca tacctcaaaa aattgaacac 360

aaatgttatc cgtgtctacg ctatcaatac cactctagat cactccgaat gtatgaaggc 420

tttgaatgat gctgacatct atgtcatcgc tgatttagca gctccagcca cctctatcaa 480

tagagacgat ccaacttgga ctgttgactt gttcaacagc tacaaaaccg ttgttgacac 540

ttttgctaat tacaccaacg ttttgggttt cttcgccggt aatgaagtta ctaacaatta 600

caccaacaca gatgcatctg ctttcgtgaa ggcagctatt agagacgtcc tggccgcctc 660

ggcc 664

<210> 41
<211> 55
<212> PRT
<213> Artificial sequence

<220>
<223> TFP21-aa

<400> 41

Met Leu Gln Ser Val Val Phe Phe Ala Leu Leu Thr Phe Ala Ser Ser
1 5 10 15

Val Ser Ala Ile Tyr Ser Asn Asn Thr Val Ser Thr Thr Thr Thr Leu
20 25 30

Ala Pro Ser Tyr Ser Leu Val Pro Gln Glu Thr Thr Ile Ser Tyr Ala
35 40 45

Asp Asp Leu Ala Ala Ser Ala
50 55

<210> 42
<211> 407
<212> DNA
<213> Artificial sequence

<220>
<223> TFP21-nt

<400> 42
ggccattacg gccggggaag caactagttt agcacaacat ccaaccaaga ggtttctcgc 60

gtatttctct catttttta cccattttac aaatttttt tgctatttga gccatagtac 120

ccattaatag gtctcgtcca ttcccttgtt ttttttttat tgtttcaatt acactacata 180

attaaaaatc acatcacttt cactctcacc ttagtcgttc tttatcaacc aaaaataaaa 240

aaatgcttca atccgttgtc tttttcgctc ttttaacctt cgcaagttct gtgtcagcga 300

tttattcaaa caatactgtt tctacaacta ccactttagc gcccagctac tccttggtgc 360

cccaagagac taccatatcg tacgccgacg acctggccgc ctcggcc 407

<210> 43
<211> 190
<212> PRT
<213> Artificial sequence

<220>
<223> TFP25-aa

<400> 43

Met Asn Trp Leu Phe Leu Val Ser Leu Val Phe Phe Cys Gly Val Ser
1 5 10 15

Thr His Pro Ala Leu Ala Met Ser Ser Asn Arg Leu Leu Lys Leu Ala
20 25 30

Asn Lys Ser Pro Lys Lys Ile Ile Pro Leu Lys Asp Ser Ser Phe Glu
35 40 45

Asn Ile Leu Ala Pro Pro His Glu Asn Ala Tyr Ile Val Ala Leu Phe
50 55 60

Thr Ala Thr Ala Pro Glu Ile Gly Cys Ser Leu Cys Leu Glu Leu Glu
65 70 75 80

Ser Glu Tyr Asp Thr Ile Val Ala Ser Trp Phe Asp Asp His Pro Asp
85 90 95

Ala Lys Ser Ser Asn Ser Asp Thr Ser Ile Phe Phe Thr Lys Val Asn
100 105 110

Leu Glu Asp Pro Ser Lys Thr Ile Pro Lys Ala Phe Gln Phe Phe Gln
115 120 125

Leu Asn Asn Val Pro Arg Leu Phe Ile Phe Lys Pro Asn Ser Pro Ser
130 135 140

Ile Leu Asp His Ser Val Ile Ser Ile Ser Thr Asp Thr Gly Ser Glu
145 150 155 160

Arg Met Lys Gln Ile Ile Gln Ala Ile Lys Gln Phe Ser Gln Val Asn
165 170 175

Asp Phe Ser Leu His Leu Pro Val Gly Leu Ala Ala Ser Ala
180 185 190

<210> 44
<211> 654
<212> DNA
<213> Artificial sequence

<220>
<223> TFP25-nt

<400> 44
ggccattacg gccgggggaa taccaggcac acgctcttcg aacactgaac cacacgcgtc 60

cgcatcaaac tcttcctccc aaacatgaat tggctgtttt tggtctcgct ggttttcttc 120

tgcggcgtgt caacccatcc tgccctggca atgtccagca acagactact aaagctggct 180

aataaatctc ccaagaaaat tatacctctg aaggactcaa gttttgaaaa catcttggca 240

ccacctcacg aaaatgccta tatagttgct ctgtttactg ccacagcgcc cgaaattggc 300

tgttctctgt gtctcgagct agaatccgaa tacgacacca tagtggcctc ctggtttgat 360

gatcatccgg atgcaaaatc gtccaattcc gatacatcta ttttcttcac aaaggtcaat 420

ttggaggacc cttctaagac cattcctaaa gcgttccagt ttttccaact aaacaatgtt 480

cctagattgt tcatcttcaa accaaactct ccctctattc tggaccacag cgtgatcagt 540

atttccactg atactggctc agaaagaatg aagcaaatca tacaagccat taagcagttc 600

tcgcaagtaa acgacttctc tttacactta cctgtgggtc tggccgcctc ggcc 654

<210> 45
<211> 89
<212> PRT
<213> Artificial sequence

<220>
<223> TFP27-aa

<400> 45

Met Lys Leu Ser Ala Leu Leu Ala Leu Ser Ala Ser Thr Ala Val Leu
1 5 10 15

Ala Ala Pro Ala Val His His Ser Asp Asn His His His Asn Asp Lys
20 25 30

Arg Ala Val Val Thr Val Thr Gln Tyr Val Asn Ala Asp Gly Ala Val
35 40 45

Val Ile Pro Ala Ala Thr Thr Ala Thr Ser Ala Ala Ala Asp Gly Lys
50 55 60

Val Glu Ser Val Ala Ala Ala Thr Thr Thr Leu Ser Ser Thr Ala Ala
65 70 75 80

Ala Ala Thr Thr Leu Ala Ala Ser Ala
85

<210> 46
<211> 470
<212> DNA
<213> Artificial sequence

<220>
<223> TFP27-nt

<400> 46
ggccattacg gccggggacg ctcctttcat cggtaactaa gaagaaaaaa aaaaaagtac 60

gaccacacaa tttccagtgt attcattcct taaacttcgt ttatttttta ttcattcatt 120

catttttatt tgaatataac caactactag tccttccttt aaacaaaaat ttaccctccc 180

ttaatttttc aagaaattcc agtatgaaat tatccgctct attagcttta tcagcctcca 240

ccgccgtctt ggccgctcca gctgtccacc atagtgacaa ccaccaccac aacgacaagc 300

gtgccgttgt caccgttact cagtacgtca acgcagacgg cgctgttgtt attccagctg 360

ccaccaccgc tacctcggcg gctgctgatg gaaaggtcga gtctgttgct gctgccacca 420

ctactttgtc ctcgactgcc gccgccgcta caaccctggc cgcctcggcc 470

<210> 47
<211> 33
<212> DNA
<213> Artificial sequence

<220>
<223> BamH-YGR-F

<400> 47
ccggatccat ggtgttcggt cagctgtatg ccc 33

<210> 48

<211> 34
<212> DNA
<213> Artificial sequence

<220>
<223> BamH-SIM-F

<400> 48
cggatccatg aaattctcaa ctgccgttac tacg 34

<210> 49
<211> 31
<212> DNA
<213> Artificial sequence

<220>
<223> BamH-YNL-F

<400> 49
ccggatccat gaagttctct tctgttactg c 31

<210> 50
<211> 31
<212> DNA
<213> Artificial sequence

<220>
<223> BamH-ECM-F

<400> 50
ccggatccat gcaattcaag aacgctttga c 31

<210> 51
<211> 34
<212> DNA
<213> Artificial sequence

<220>
<223> BamH-ATG-F

<400> 51
ccggatccat ggtatcgaag acttggatat gtgg 34

<210> 52
<211> 37
<212> DNA
<213> Artificial sequence

<220>
<223> BamH-GAS-F

<400> 52
ccggatccat gttgtttaaa tccctttcaa agttagc 37

<210> 53
<211> 34

<212> DNA
<213> Artificial sequence

<220>
<223> BamH-YOR-F

<400> 53
ccggatccat gcttcaatcc gttgtctttt tcgc 34

<210> 54
<211> 36
<212> DNA
<213> Artificial sequence

<220>
<223> BamH-OST-F

<400> 54
ccggatccat gaattggctg tttttggtct cgctgg 36

<210> 55
<211> 32
<212> DNA
<213> Artificial sequence

<220>
<223> BamH-UTH-F

<400> 55
ccggatccat gtgtttcctt ctcgagacct cg 32

<210> 56
<211> 43
<212> DNA
<213> Artificial sequence

<220>
<223> IL2-TGA-R

<400> 56
gtcactccgt tcaagtcgac tcaagttagt gttgagatga tgc 43

<210> 57
<211> 24
<212> DNA
<213> Artificial sequence

<220>
<223> Sac-GAL-F

<400> 57
gagctcatcg cttcgctgat taat 24

<210> 58
<211> 27
<212> DNA

<213> Artificial sequence

<220>
<223> GAL-BamH-R

<400> 58
ggatcctgaa ttttcaaaaa ttcttac 27

<210> 59
<211> 38
<212> DNA
<213> Artificial sequence

<220>
<223> KR-IL32Alpha-F

<400> 59
ctcgccttag ataaaagaat gtgcttcccg aaggtcct 38

<210> 60
<211> 38
<212> DNA
<213> Artificial sequence

<220>
<223> IL32-INV-R

<400> 60
ctcgccttag ataaaagaat gtgcttcccg aaggtcct 38

<210> 61
<211> 187
<212> PRT
<213> Artificial sequence

<220>
<223> TFP-11-AA

<400> 61

Met Lys Leu Ser Thr Val Leu Leu Ser Ala Gly Leu Ala Ser Thr Thr
1 5 10 15

Leu Ala Gln Phe Ser Asn Ser Thr Ser Ala Ser Ser Thr Asp Val Thr
20 25 30

Ser Ser Ser Ser Ile Ser Thr Ser Ser Gly Ser Val Thr Ile Thr Ser
35 40 45

Ser Glu Ala Pro Glu Ser Asp Asn Gly Thr Ser Thr Ala Ala Pro Thr
50 55 60

Glu Thr Ser Thr Glu Ala Pro Thr Thr Ala Ile Pro Thr Asn Gly Thr
65 70 75 80

Ser Thr Glu Ala Pro Thr Thr Ala Ile Pro Thr Asn Gly Thr Ser Thr
85 90 95

Glu Ala Pro Thr Asp Thr Thr Thr Glu Ala Pro Thr Thr Ala Leu Pro
100 105 110

Thr Asn Gly Thr Ser Thr Glu Ala Pro Thr Asp Thr Thr Thr Glu Ala
115 120 125

Pro Thr Thr Gly Leu Pro Thr Asn Gly Thr Thr Ser Ala Phe Pro Pro
130 135 140

Thr Thr Ser Leu Pro Pro Ser Asn Thr Thr Thr Thr Pro Pro Tyr Asn
145 150 155 160

Pro Ser Thr Asp Tyr Thr Thr Asp Tyr Thr Val Val Thr Glu Tyr Thr
165 170 175

Thr Tyr Cys Pro Glu Pro Leu Ala Ala Ser Ala
180 185

<210> 62
<211> 621
<212> DNA
<213> Artificial sequence

<220>
<223> TFP-11-NT

<400> 62
ggccattacg gccggggagc tacaaagaca agcaaaataa aatacgttcg ctctattaag 60

atgaaattat caactgtcct attatctgcc ggtttagcct cgactacttt ggcccaattt 120

tccaacagta catctgcttc ttccaccgat gtcacttcct cctcttccat ctccacttcc 180

tctggctcag taactatcac atcttctgaa gctccagaat ccgacaacgg taccagcaca 240

gctgcaccaa ctgaaacctc aacagaggct ccaaccactg ctatcccaac taacggtacc 300

tctactgaag ctccaaccac tgctatccca actaacggta cctctactga agctccaact 360

gatactacta ctgaagctcc aaccaccgct cttccaacta acggtacttc tactgaagct 420

ccaactgata ctactactga agctccaacc accggtcttc caaccaacgg taccacttca 480

gctttcccac caactacatc tttgccacca agcaacacta ccaccactcc tccttacaac 540

ccatctactg actacaccac tgactacact gtagtcactg aatatactac ttactgtccg 600

gaaccactgg ccgcctcggc c 621

<210> 63
<211> 165
<212> PRT
<213> Artificial sequence

<220>
<223> TFP-22-AA

<400> 63

Met Gln Tyr Lys Lys Thr Leu Val Ala Ser Ala Leu Ala Ala Thr Thr
1 5 10 15

Leu Ala Ala Tyr Ala Pro Ser Glu Pro Trp Ser Thr Leu Thr Pro Thr
20 25 30

Ala Thr Tyr Ser Gly Gly Val Thr Asp Tyr Ala Ser Thr Phe Gly Ile
35 40 45

Ala Val Gln Pro Ile Ser Thr Thr Ser Ser Ala Ser Ser Ala Ala Thr
50 55 60

Thr Ala Ser Ser Lys Ala Lys Arg Ala Ala Ser Gln Ile Gly Asp Gly
65 70 75 80

Gln Val Gln Ala Ala Thr Thr Thr Ala Ser Val Ser Thr Lys Ser Thr
85 90 95

Ala Ala Ala Val Ser Gln Ile Gly Asp Gly Gln Ile Gln Ala Thr Thr
100 105 110

Lys Thr Thr Ala Ala Ala Val Ser Gln Ile Gly Asp Gly Gln Ile Gln
115 120 125

Ala Thr Thr Lys Thr Thr Ser Ala Lys Thr Thr Ala Ala Ala Val Ser
130 135 140

Gln Ile Ser Asp Gly Gln Ile Gln Ala Thr Thr Thr Thr Leu Ala Pro
145 150 155 160

Leu Ala Ala Ser Ala
165

<210> 64
<211> 564
<212> DNA
<213> Artificial sequence

<220>

<223> TFP-22-NT

<400> 64
ggccattacg gccgggggaa taagaaactc atattccttt tctaacccta gtacaataat 60

aataatataa tgcaatacaa aaagactttg gttgcctctg ctttggccgc tactacattg 120

gccgcctatg ctccatctga gccttggtcc actttgactc caacagccac ttacagcggt 180

ggtgttaccg actacgcttc caccttcggt attgccgttc aaccaatctc cactacatcc 240

agcgcatcat ctgcagccac cacagcctca tctaaggcca agagagctgc ttcccaaatt 300

ggtgatggtc aagtccaagc tgctaccact actgcttctg tctctaccaa gagtaccgct 360

gccgccgttt ctcagatcgg tgatggtcaa atccaagcta ctaccaagac taccgctgct 420

gctgtctctc aaattggtga tggtcaaatt caagctacca ccaagactac ctctgctaag 480

actaccgccg ctgccgtttc tcaaatcagt gatggtcaaa tccaagctac caccactact 540

ttagcccctc tggccgcctc ggcc 564

<210> 65
<211> 48
<212> PRT
<213> Artificial sequence

<220>
<223> TFP-29-AA

<400> 65

Met Lys Leu Glu Asn Thr Leu Phe Thr Leu Gly Ala Leu Gly Ser Ile
1 5 10 15

Ser Ala Ala Leu Val Ile Pro Asn Leu Glu Asn Ala Ala Asp His His
20 25 30

Glu Leu Ile Asn Lys Glu Asp His His Glu Arg Leu Ala Ala Ser Ala
35 40 45

<210> 66
<211> 216
<212> DNA
<213> Artificial sequence

<220>
<223> TFP-29-NT

<400> 66
ggccattacg gccggggaat tagcttcatc gccaataaaa aaacaaacta aacctaattc 60

taacaagcaa agatgaagtt agaaaatact ctatttacac tcggtgccct agggagcatc 120

tctgctgctt tggtcatccc aaatcttgaa aatgccgccg accaccacga actgattaac 180

aaggaagatc accacgagag actggccgcc tcggcc 216

<210> 67
<211> 208
<212> PRT
<213> Artificial sequence

<220>
<223> TFP-34-AA

<400> 67

Met Arg Ala Thr Thr Leu Leu Ser Ser Val Val Ser Leu Ala Leu Leu
1 5 10 15

Ser Lys Glu Val Leu Ala Thr Pro Pro Ala Cys Leu Leu Ala Cys Val
20 25 30

Ala Gln Val Gly Lys Ser Ser Ser Thr Cys Asp Ser Leu Asn Gln Val
35 40 45

Thr Cys Tyr Cys Glu His Glu Asn Ser Ala Val Lys Lys Cys Leu Asp
50 55 60

Ser Ile Cys Pro Asn Asn Asp Ala Asp Ala Ala Tyr Ser Ala Phe Lys
65 70 75 80

Ser Ser Cys Ser Glu Gln Asn Ala Ser Leu Gly Asp Ser Ser Ser Ser
85 90 95

Ala Ser Ser Ser Ala Ser Ser Ser Ser Lys Ala Ser Ser Ser Thr Lys
100 105 110

Ala Ser Ser Ser Ser Ala Ser Ser Ser Thr Lys Ala Ser Ser Ser Ser
115 120 125

Ala Ser Ser Pro Thr Lys Ala Ser Ser Ser Ser Ala Ala Pro Ser Ser
130 135 140

Ser Lys Ala Ser Ser Thr Glu Ser Ser Ser Ser Ser Ser Ser Ser Thr
145 150 155 160

Lys Ala Pro Ser Ser Glu Glu Ser Ser Ser Thr Tyr Val Ser Ser Ser
165 170 175

Lys Gln Ala Ser Ser Thr Ser Glu Ala His Ser Ser Ser Ala Ala Ser
180 185 190

Ser Thr Val Ser Gln Glu Thr Val Ser Ser Ala Leu Ala Ala Ser Ala
195 200 205

<210> 68
<211> 694
<212> DNA
<213> Artificial sequence

<220>
<223> TFP-34-NT

<400> 68
ggccattacg gccgggaggg tcaaagctca cagcactact acactcgttc aacactcgtt 60

atatattatc atgcgcgcca ccactttatt atcttcagtc gtttctttgg cattgttgtc 120

gaaggaagtc ttagcaacac ctccagcttg tttattggcc tgtgttgcgc aagtcggcaa 180

atcctcttcc acatgtgact ctttgaatca agtcacctgt tactgtgaac acgaaaactc 240

cgccgtcaag aaatgtctag actccatctg cccaaacaat gacgctgatg ctgcttattc 300

tgctttcaag agttcttgtt ccgaacaaaa tgcttcattg ggcgattcca gcagcagtgc 360

ctcctcatcc gcttcttcat ccagcaaggc ctcttcttct accaaggctt cttccagtag 420

cgcttcctcc tctaccaagg cttcttccag tagcgcttcc tcccctacta aagcttcttc 480

cagcagcgct gccccatctt ctagcaaggc ttcttccacc gaatcctctt cttcctcttc 540

ttcttccacc aaggctcctt ccagtgaaga atcctcttcc acttatgtct cttcgagcaa 600

gcaagcttcc tccactagcg aggctcactc ttccagtgct gcctcttcga ccgtgtccca 660

agaaacagtc tcctctgctc tggccgcctc ggcc 694

<210> 69
<211> 38
<212> PRT
<213> Artificial sequence

<220>
<223> TFP-38-AA

<400> 69

Met Lys Leu Ser Gln Val Val Val Ser Ala Val Ala Phe Thr Gly Leu
1 5 10 15

Val Ser Ala Ala Asn Ser Ser Asn Ser Ser Ser Ser Lys Asn Ala Ala
20 25 30

Gln Leu Ala Ala Ser Ala
35

<210> 70

<211> 184
<212> DNA
<213> Artificial sequence

<220>
<223> TFP-38-NT

<400> 70
ggccattacg gccgggggac tatcaaatca tacagatatt gtcaaaaaaa aaaaagacta 60

ataataaaaa atgaagttat ctcaagttgt tgtttccgcc gtcgccttca ctggtttagt 120

aagtgctgct aacagttcta acagctcaag ctcaaagaat gctgcccaac tggccgcctc 180

ggcc 184

<210> 71
<211> 26
<212> DNA
<213> Artificial sequence

<220>
<223> BamH-CIS-F

<400> 71
ccggatccat gcaattcaaa aacgtc 26

<210> 72
<211> 37
<212> DNA
<213> Artificial sequence

<220>
<223> BamH-SED-F

<400> 72
ccggatccat gaaattatca actgtcctat tatctgc 37

<210> 73
<211> 35
<212> DNA
<213> Artificial sequence

<220>
<223> BamH-SIM-F

<400> 73
ccggatccat gaaattctca actgccgtta ctacg 35

<210> 74
<211> 34
<212> DNA
<213> Artificial sequence

<220>
<223> BamH-YOR247W-F

<400> 74
ccggatccat gcttcaatcc gttgtctttt tcgc 34

<210> 75
<211> 30
<212> DNA
<213> Artificial sequence

<220>
<223> BamH-HSP-F

<400> 75
ccggatccat gcaatacaaa aagactttgg 30

<210> 76
<211> 36
<212> DNA
<213> Artificial sequence

<220>
<223> BamH-OST-F

<400> 76
ccggatccat gaattggctg tttttggtct cgctgg 36

<210> 77
<211> 41
<212> DNA
<213> Artificial sequence

<220>
<223> IL32-TGA-R

<400> 77
cactccgttc aagtcgactc attttgagga ttggggttca g 41

<210> 78
<211> 30
<212> DNA
<213> Artificial sequence

<220>
<223> H77-1-R

<400> 78
aagtcgacat ttaaatcttt tatctaaggc 30

<210> 79
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<223> hGH-F

<400> 79

ttcccaacca ttcccttatc 20

<210> 80
<211> 19
<212> DNA
<213> Artificial sequence

<220>
<223> hGH-R

<400> 80
ctagaagcca cagctgccc 19

<210> 81
<211> 38
<212> DNA
<213> Artificial sequence

<220>
<223> KR-hGH-F

<400> 81
ctcgccttag ataaaagatt cccaaccatt cccttatc 38

<210> 82
<211> 37
<212> DNA
<213> Artificial sequence

<220>
<223> hGH-Sal-R

<400> 82
cactccgttc aagtcgacct agaagccaca gctgccc 37

<210> 83
<211> 70
<212> DNA
<213> Artificial sequence

<220>
<223> GT70-R

<400> 83
tcagatttac agataatgat gtcattatta aatatatata tatatatatt gtcactccgt 60

tcaagtcgac 70

<210> 84
<211> 101
<212> PRT
<213> Artificial sequence

<220>
<223> PpTFP1-aa

<400> 84

Met Gln Phe Asn Ser Val Val Ile Ser Gln Leu Leu Leu Thr Leu Ala
1 5 10 15

Ser Val Ser Met Gly Ala Ser Thr Ala Phe Lys Glu His His Gln His
20 25 30

Gln Arg Ala Thr Leu Glu Lys Arg Ala Thr Thr Cys Lys Phe Pro Thr
35 40 45

Asp Lys Asn Leu Val Ala Val Thr Pro Asn Ser Lys Asn Gly Gly Trp
50 55 60

Ala Leu Ser Pro Asp Gln Glu Cys Thr Ala Gly Ser Tyr Cys Pro Tyr
65 70 75 80

Ala Cys Pro Pro Gly Gln Leu Met Ala Gln Trp Asp Pro Ser Ala Thr
85 90 95

Leu Ala Ala Ser Ala
100

<210> 85
<211> 419
<212> DNA
<213> Artificial sequence

<220>
<223> PpTFP1-nt

<400> 85
ggccattacg ccggggcaca gtaactttga cataatatct ggtagctgca tcacttcacc 60

gactattcat tccttccttt ttagtattac caactatatc acattccttt aagaaaatgc 120

aattcaacag tgtcgtcatc agccaacttt tgctgactct agccagtgtc tcaatgggag 180

cttcaaccgc tttcaaggag caccaccagc accaaagagc tactctagag aagagagcta 240

ctacctgcaa attccccact gacaaaaact tggtcgctgt tacaccaaac tccaaaaatg 300

gaggctgggc tctgagtcca gaccaggagt gcacagcagg ttcttattgt ccttatgctt 360

gtccaccagg ccagttgatg gctcaatggg acccatcggc cacactggcc gcctcggcc 419

<210> 86
<211> 94
<212> PRT
<213> Artificial sequence

<220>
<223> PpTFP2-aa

<400> 86

Met Gln Phe Ser Ile Val Ala Thr Leu Ala Leu Ala Gly Ser Ala Leu
1 5 10 15

Ala Ala Tyr Ser Asn Val Thr Tyr Thr Tyr Glu Thr Thr Ile Thr Asp
20 25 30

Val Val Thr Glu Leu Thr Thr Tyr Cys Pro Glu Pro Thr Thr Phe Val
35 40 45

His Lys Asn Lys Thr Ile Thr Val Thr Ala Pro Thr Thr Leu Thr Ile
50 55 60

Thr Asp Cys Pro Cys Thr Ile Ser Lys Thr Thr Lys Ile Thr Thr Asp
65 70 75 80

Val Pro Pro Thr Thr His Ser Thr Pro Leu Ala Ala Ser Ala
85 90

<210> 87
<211> 345
<212> DNA
<213> Artificial sequence

<220>
<223> PpTFP2-nt

<400> 87
ggccattacg gccgggggac ttacatttta ccgttccgtc actcgcttca ctcaacaaca 60

aaaatgcaat tctctatcgt cgctactttg gctcttgctg gttccgctct ggctgcttac 120

tctaacgtaa cttacactta cgagactacc atcaccgatg ttgtcaccga gctcaccact 180

tactgcccag agccaaccac cttcgttcac aagaacaaga ccatcactgt gaccgcccca 240

accactttga ccatcactga ctgtccttgc accatctcca agaccaccaa gatcaccact 300

gatgttccac caaccaccca ctccacccca ctggccgcct cggcc 345

<210> 88
<211> 82
<212> PRT
<213> Artificial sequence

<220>
<223> PpTFP3-aa

<400> 88

Met Lys Phe Ser Thr Ala Phe Ala Gly Phe Val Ala Leu Asn Ala Val
1 5 10 15

Ser Ile Val Ala Gln Asp Glu Ala Thr Asp Ala His Val Val Thr Thr
20 25 30

Thr Val Thr Thr Ala Ser Thr Glu Thr His Arg Trp Gly Arg Phe Asp
35 40 45

Lys Thr Ser Pro Pro Thr Thr Ser Thr Ser Ser Gly Thr His Arg Trp
50 55 60

Gly Arg Phe Asn Lys Thr Pro Asp Pro Thr Thr Thr Ser Ala Ala
65 70 75 80

Ser Ala

<210> 89
<211> 273
<212> DNA
<213> Artificial sequence

<220>
<223> PpTFP3-nt

<400> 89
ggccattacg gccggggaac atcaagaatg aagttttcca ctgcgtttgc tggctttgtt 60

gccctaaatg ctgtgtccat tgttgctcag gacgaggcta ccgatgctca cgttgtcacc 120

acaactgtga ccaccgcttc cactgagact cacagatggg gaagattcga caagacttct 180

cctcctacaa cttccacttc ttcaggtact cacagatggg gaagatttaa caaaactcca 240

gatcctacca ctaccacctc ggccgcctcg gcc 273

<210> 90
<211> 127
<212> PRT
<213> Artificial sequence

<220>
<223> PpTFP4-aa

<400> 90

Met Gln Tyr Arg Ser Leu Phe Leu Gly Ser Ala Leu Leu Ala Ala Ala
1 5 10 15

Asn Ala Ala Val Tyr Asn Thr Thr Val Thr Asp Val Val Ser Glu Leu
20 25 30

Glu Thr Thr Val Leu Thr Ile Thr Ser Cys Ala Glu Asp Lys Cys Ile

35 40 45

Thr Ser Lys Ser Thr Gly Leu Ile Thr Thr Ser Thr Leu Thr Lys His
50 55 60

Gly Val Val Thr Val Val Thr Thr Val Cys Asp Leu Pro Ser Thr Thr
65 70 75 80

Lys Ser Tyr Val Pro Pro Ala Lys Thr Thr Thr Ile Pro Pro Pro Glu
85 90 95

Lys Thr Thr Thr Thr Val Pro Pro Pro Ala Lys Thr Thr Thr Thr Val
100 105 110

Pro Pro Pro Ala Lys Thr Thr Ser Thr Ala Leu Ala Ala Ser Ala
115 120 125

<210> 91
<211> 444
<212> DNA
<213> Artificial sequence

<220>
<223> PpTFP4-nt

<400> 91
ggccattacg gggggaactc actgtttcag tttattccaa ctactttcac tcacttatca 60

aaaatgcaat acagatctct ctttttaggt tccgccttat tggccgctgc taacgctgct 120

gtttacaaca ccaccgtcac tgacgttgtt tccgagttgg agaccaccgt tctgactatc 180

acctcttgtg ctgaggacaa gtgtatcacc agtaagtcca ccggattgat cactacctcc 240

accctcacca agcacggtgt tgtcactgtt gtcaccactg tctgtgactt gccaagcacc 300

accaagagct acgtcccacc tgctaagact actactattc ctcctccaga gaagactacc 360

accactgtcc cacctccagc caagactacc accactgtcc cacctccagc caagactact 420

agtaccgccc tggccgcctc ggcc 444

<210> 92
<211> 38
<212> DNA
<213> Artificial sequence

<220>
<223> YGR279C-F

<400> 92
ggccattacg gccaaaatgc gtctctctaa cctaattg 38

<210> 93
<211> 21
<212> DNA
<213> Artificial sequence

<220>
<223> YGR279C-R

<400> 93
tcattggata gaatacccca g 21

<210> 94
<211> 38
<212> DNA
<213> Artificial sequence

<220>
<223> YLR037C-F

<400> 94
ggccattacg gccaaaatgg tcaaactaac ttcaattg 38

<210> 95
<211> 22
<212> DNA
<213> Artificial sequence

<220>
<223> YLR037C-R

<400> 95
ttagtttgga acagcagtgt ag 22

<210> 96
<211> 36
<212> DNA
<213> Artificial sequence

<220>
<223> YLR110C-F

<400> 96
ggccattacg gccaaaatgc aattttctac tgtcgc 36

<210> 97
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<223> YLR110C-R

<400> 97
ttacaacaac aaagcagcgg 20

<210> 98

<211> 37
<212> DNA
<213> Artificial sequence

<220>
<223> YOR383C-F

<400> 98
ggccattacg gccaaaatga aattctcttc cgctttg 37

<210> 99
<211> 21
<212> DNA
<213> Artificial sequence

<220>
<223> YOR383C-R

<400> 99
ttacaataac atgacggcag c 21

<210> 100
<211> 36
<212> DNA
<213> Artificial sequence

<220>
<223> YIL011W-F

<400> 100
ggccattacg gccaaaatgt ctttcactaa aatcgc 36

<210> 101
<211> 21
<212> DNA
<213> Artificial sequence

<220>
<223> YIL011W-R

<400> 101
tcataagagc atagcagcgg c 21

<210> 102
<211> 38
<212> DNA
<213> Artificial sequence

<220>
<223> YHR214W-F

<400> 102
ggccattacg gccaaaatgt tcaatcgttt taacaaat 38

<210> 103
<211> 22

<212> DNA
<213> Artificial sequence

<220>
<223> YHR214W-R

<400> 103
ttacaaaccg gaaacagaac ca 22

<210> 104
<211> 39
<212> DNA
<213> Artificial sequence

<220>
<223> YNL160W-F

<400> 104
ggccattacg gccaaaatga agttccaagt tgttttatc 39

<210> 105
<211> 21
<212> DNA
<213> Artificial sequence

<220>
<223> YNL160W-R

<400> 105
tcatgggaaa atgctttcca g 21

<210> 106
<211> 37
<212> DNA
<213> Artificial sequence

<220>
<223> YGR296C-A-F

<400> 106
ggccattacg gccaaaatgg aatctattat cctcagc 37

<210> 107
<211> 21
<212> DNA
<213> Artificial sequence

<220>
<223> YGR296C-A-R

<400> 107
ttaccgtcta gcttccagga g 21

<210> 108
<211> 39
<212> DNA

<213> Artificial sequence

<220>
<223> YOL154W-F

<400> 108
ggccattacg gccaaaatga agttctcttc cggcaaatc 39

<210> 109
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<223> YOL154W-R

<400> 109
agttacctag acagccacca 20

<210> 110
<211> 39
<212> DNA
<213> Artificial sequence

<220>
<223> YPL187W-F

<400> 110
ggccattacg gccaaaatga gatttccttc aatttttac 39

<210> 111
<211> 19
<212> DNA
<213> Artificial sequence

<220>
<223> YPL187W-R

<400> 111
ttagtacatt ggttggccg 19

<210> 112
<211> 34
<212> DNA
<213> Artificial sequence

<220>
<223> YHR214W-F

<400> 112
ggccattacg gccaaaatgt tcaatcgttt taac 34

<210> 113
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<223> YHR214W-

<400> 113
cggaaacaga accaccgttg 20

<210> 114
<211> 34
<212> DNA
<213> Artificial sequence

<220>
<223> YKR013W-F

<400> 114
ggccattacg gccaaaatga aattttctaa agtc 34

<210> 115
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<223> YKR013W-R

<400> 115
ctcaccaatg acattaccag 20

<210> 116
<211> 34
<212> DNA
<213> Artificial sequence

<220>
<223> YHR139C-F

<400> 116
ggccattacg gccaaaatga aattcacatc agtg 34

<210> 117
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<223> YHR139C-R

<400> 117
gtaactcgct actacttgtg 20

<210> 118
<211> 34
<212> DNA
<213> Artificial sequence

<220>
<223> YIL169C-F

<400> 118
ggccattacg gccaaaatgt tcaatcgttt aaac 34

<210> 119
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<223> YIL169C-R

<400> 119
agttgcgctt gcactagatg 20

<210> 120
<211> 34
<212> DNA
<213> Artificial sequence

<220>
<223> YOL155C-F

<400> 120
ggccattacg gccaaaatgt tcaatcgctt taat 34

<210> 121
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<223> YOL155C-R

<400> 121
agaggcagtg gaagccgatg 20

<210> 122
<211> 34
<212> DNA
<213> Artificial sequence

<220>
<223> YMR325W-F

<400> 122
ggccattacg gccaaaatgg tcaaattaac ttca 34

<210> 123
<211> 20
<212> DNA
<213> Artificial sequence

<220>

<223> YMR325W-R

<400> 123
atagcagtgt agataccgtc 20

<210> 124
<211> 34
<212> DNA
<213> Artificial sequence

<220>
<223> YDR134W-F

<400> 124
ggccattacg gccaaaatgc aattctctac cgtc 34

<210> 125
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<223> YDR134W-R

<400> 125
ttacaacaat aaagcggcag 20

<210> 126
<211> 34
<212> DNA
<213> Artificial sequence

<220>
<223> YLR300W-F

<400> 126
ggccattacg gccaaaatgc tttcgcttaa aacg 34

<210> 127
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<223> YLR300W-R

<400> 127
tgatgatggt cgatagtgac 20

<210> 128
<211> 30
<212> DNA
<213> Artificial sequence

<220>
<223> SfiA-F

<400> 128
ctgagtctca cggccattat ggccaaaatg 30

<210> 129
<211> 57
<212> PRT
<213> Artificial sequence

<220>
<223> TFP-39-AA

<400> 129

Met Arg Leu Ser Asn Leu Ile Ala Ser Ala Ser Leu Leu Ser Ala Ala
1 5 10 15

Thr Leu Ala Ala Pro Ala Asn His Glu His Lys Asp Lys Arg Ala Val
20 25 30

Val Thr Thr Thr Val Gln Lys Gln Thr Thr Val Ile Val Asn Gly Ala
35 40 45

Ala Ser Thr Pro Leu Ala Ala Ser Ala
50 55

<210> 130
<211> 214
<212> DNA
<213> Artificial sequence

<220>
<223> TFP-39-nt

<400> 130
ggccattacg gccaaaatgc gtctctctaa cctaattgct tctgcctctc ttttatctgc 60

tgctactctt gctgcccccg ctaaccacga acacaaggac aagcgtgctg tggtcactac 120

cactgttcaa aaacaaacca ctgtcattgt taatggtgcc gcttcaactc ccctggccgc 180

ctcggcctct gctggcctcg ccttagataa aaga 214

<210> 131
<211> 128
<212> PRT
<213> Artificial sequence

<220>
<223> TFP-43-AA

<400> 131

Met Gln Phe Ser Thr Val Ala Ser Ile Ala Ala Val Ala Ala Val Ala
1 5 10 15

Ser Ala Ala Ala Asn Val Thr Thr Ala Thr Val Ser Gln Glu Ser Thr
20 25 30

Thr Leu Val Thr Ile Thr Ser Cys Glu Asp His Val Cys Ser Glu Thr
35 40 45

Val Ser Pro Ala Leu Val Ser Thr Ala Thr Val Thr Val Asp Asp Val
50 55 60

Ile Thr Gln Tyr Thr Thr Trp Cys Pro Leu Thr Thr Glu Ala Pro Lys
65 70 75 80

Asn Gly Thr Ser Thr Ala Ala Pro Val Thr Ser Thr Glu Ala Pro Lys
85 90 95

Asn Thr Thr Ser Ala Ala Pro Thr His Ser Val Thr Ser Tyr Thr Gly
100 105 110

Ala Ala Ala Lys Ala Leu Pro Ala Ala Gly Ala Leu Leu Ala Ala Ser
115 120 125

<210> 132
<211> 403
<212> DNA
<213> Artificial sequence

<220>
<223> TFP-43-nt

<400> 132
ggccattacg gccaaaatgc aattttctac tgtcgcttct atcgccgctg tcgccgctgt 60

cgcttctgcc gctgctaacg ttaccactgc tactgtcagc caagaatcta ccactttggt 120

caccatcact tcttgtgaag accacgtctg ttctgaaact gtctccccag ctttggtttc 180

caccgctacc gtcaccgtcg atgacgttat cactcaatac accacctggt gcccattgac 240

cactgaagcc ccaaagaacg gtacttctac tgctgctcca gttacctcta ctgaagctcc 300

aaagaacacc acctctgctg ctccaactca ctctgtcacc tcttacactg gtgctgctgc 360

taaggctttg ccagctgctg gtgctttgct ggccgcctcg gcc 403

<210> 133
<211> 71
<212> PRT
<213> Artificial sequence

<220>
<223> TFP-44-AA

<400> 133

Met Lys Phe Ser Ser Ala Leu Val Leu Ser Ala Val Ala Ala Thr Ala
1 5 10 15

Leu Ala Glu Ser Ile Thr Thr Thr Ile Thr Ala Thr Lys Asn Gly His
20 25 30

Val Tyr Thr Lys Thr Val Thr Gln Asp Ala Thr Phe Val Trp Gly Gly
35 40 45

Glu Asp Ser Tyr Ala Ser Ser Thr Ser Ala Ala Glu Ser Ser Ala Ala
50 55 60

Glu Thr Ser Ala Ala Ser Ala
65 70

<210> 134
<211> 229
<212> DNA
<213> Artificial sequence

<220>
<223> TFP-44-nt

<400> 134
ggccattacg gccaaaatga aattctcttc cgctttggtt ctatctgctg ttgccgctac 60

tgctcttgct gagagtatca ccaccaccat cactgccacc aagaacggtc atgtctacac 120

taagactgtc acccaagatg ctacttttgt ttggggtggt gaagactctt acgccagcag 180

cacttctgcc gctgaatctt ctgccgccga aacttcggcc gcctcggcc 229

<210> 135
<211> 119
<212> PRT
<213> Artificial sequence

<220>
<223> TFP-48-AA

<400> 135

Met Arg Leu Ser Asn Leu Ile Ala Ser Ala Ser Leu Leu Ser Ala Ala
1 5 10 15

Thr Leu Ala Ala Pro Ala Asn His Glu His Lys Asp Lys Arg Ala Val
20 25 30

Val Thr Thr Thr Val Gln Lys Gln Thr Thr Ile Ile Val Asn Gly Ala
35 40 45

Ala Ser Thr Pro Val Ala Ala Leu Glu Glu Asn Ala Val Val Asn Ser
50 55 60

Ala Pro Ala Ala Ala Thr Ser Thr Thr Ser Ser Ala Ala Ser Val Ala
65 70 75 80

Thr Ala Ala Ala Ser Ser Ser Glu Asn Asn Ser Gln Val Ser Ala Ala
85 90 95

Ala Ser Pro Ala Ser Ser Ser Ala Ala Thr Ser Thr Gln Ser Ser Ser
100 105 110

Ser Ser Leu Ala Ala Ser Ala
115

<210> 136
<211> 373
<212> DNA
<213> Artificial sequence

<220>
<223> TFP-48-nt

<400> 136
ggccattacg gccaaaatgc gtctctctaa cctaattgct tctgcctctc ttttatctgc 60

tgctactctt gctgctcccg ctaaccacga acacaaggac aagcgtgctg tggtcactac 120

cactgttcaa aaacaaacca ctatcattgt taatggtgcc gcttcaactc cagttgctgc 180

tttggaagaa aatgctgttg tcaactccgc tccagctgcc gctaccagta caacatcgtc 240

tgctgcttct gtagctaccg ctgctgcttc ctcttctgag aacaactcac aagtttctgc 300

tgccgcatct ccagcctcca gctctgctgc tacatctact caatcttcct cttcctccct 360

ggccgcctcg gcc 373

<210> 137
<211> 129
<212> PRT
<213> Artificial sequence

<220>
<223> TFP-52-AA

<400> 137

Met Lys Phe Gln Val Val Leu Ser Ala Leu Leu Ala Cys Ser Ser Ala
1 5 10 15

Val Val Ala Ser Pro Ile Glu Asn Leu Phe Lys Tyr Arg Ala Val Lys

20 25 30

Ala Ser His Ser Lys Asn Ile Asn Ser Thr Leu Pro Ala Trp Asn Gly
35 40 45

Ser Asn Ser Ser Asn Val Thr Tyr Ala Asn Gly Thr Asn Ser Thr Thr
50 55 60

Asn Thr Thr Thr Ala Glu Ser Ser Gln Leu Gln Ile Ile Val Thr Gly
65 70 75 80

Gly Gln Val Pro Ile Thr Asn Ser Ser Leu Thr His Thr Asn Tyr Thr
85 90 95

Arg Leu Phe Asn Ser Ser Ser Ala Leu Asn Ile Thr Glu Leu Tyr Asn
100 105 110

Val Ala Arg Val Val Asn Glu Thr Ile Gln Asp Asn Leu Ala Ala Ser
115 120 125

Ala

<210> 138
<211> 403
<212> DNA
<213> Artificial sequence

<220>
<223> TFP-52-nt

<400> 138
ggccattacg gccaaaatga agttccaagt tgttttatct gccctttgg catgttcatc 60

tgccgtcgtc gcaagcccaa tcgaaaacct attcaaatac agggcagtta aggcatctca 120

cagtaagaat atcaactcca ctttgccggc ctggaatggg tctaactcta gcaatgttac 180

ctacgctaat ggaacaaaca gtactaccaa tactactact gccgaaagca gtcaattaca 240

aatcattgta acaggtggtc aagtaccaat caccaacagt tctttgaccc acacaaacta 300

caccagatta ttcaacagtt cttctgcttt gaacattacc gaattgtaca atgttgcccg 360

tgttgttaac gaaacgatcc aagataacct ggccgcctcg gcc 403

<210> 139
<211> 124
<212> PRT
<213> Artificial sequence

<220>

<223> TFP-54-aa

<400> 139

Met Val Lys Leu Thr Ser Ile Val Ala Gly Val Ala Ala Ile Ala Ala
1 5 10 15

Gly Val Ala Ala Ala Pro Ala Thr Thr Thr Leu Ser Pro Ser Asp Glu
20 25 30

Arg Val Asn Leu Val Glu Leu Gly Val Tyr Val Ser Asp Ile Arg Ala
35 40 45

His Leu Ala Glu Tyr Tyr Met Phe Gln Ala Ala His Pro Thr Glu Thr
50 55 60

Tyr Pro Val Glu Ile Ala Glu Ala Val Phe Asn Tyr Gly Asp Phe Thr
65 70 75 80

Thr Met Leu Thr Gly Ile Pro Ala Asp Gln Val Thr Arg Val Ile Thr
85 90 95

Gly Val Pro Trp Tyr Ser Thr Arg Leu Arg Pro Ala Ile Ser Ser Ala
100 105 110

Leu Ser Lys Asp Gly Ile Tyr Thr Ala Ala Ser Ala
115 120

<210> 140
<211> 388
<212> DNA
<213> Artificial sequence

<220>
<223> TFP-54-nt

<400> 140
ggccattacg gccaaaatgg tcaaactaac ttcaattgtt gctggtgtcg ctgctattgc 60

tgctggtgtc gctgctgccc cagccaccac tactttatct ccctctgatg aaagagttaa 120

cctggtcgaa ttaggtgtct acgtctcaga tatcagagct catttggctg aatactatat 180

gttccaagct gctcatccaa ctgaaactta cccagttgaa attgctgaag ctgttttcaa 240

ctacggtgat ttcaccacta tgttgactgg tattcccgct gatcaagtca ctagagtcat 300

cactggtgtc ccatggtact ccaccagatt gagaccagct atctccagcg ctctatccaa 360

ggacggtatc tacacggccg cctcggcc 388

<210> 141

<211> 35
<212> DNA
<213> Artificial sequence

<220>
<223> YAR066W-F

<400> 141
ggccattatg gccaaaatgt tcaatcgttt taaca 35

<210> 142
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<223> YAR066W-R

<400> 142
gaaccaccgt tgagaatagc 20

<210> 143
<211> 35
<212> DNA
<213> Artificial sequence

<220>
<223> YFR026C-F(

<400> 143
ggccattatg gccaaaatga cgccctatgc agtag 35

<210> 144
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<223> YFR026C-R

<400> 144
tcactttcca gagctataag 20

<210> 145
<211> 35
<212> DNA
<213> Artificial sequence

<220>
<223> YJL158C-F

<400> 145
ggccattatg gccaaaatgc aattcaaaaa cgtcg 35

<210> 146
<211> 21

<212> DNA
<213> Artificial sequence

<220>
<223> YJL158C-R

<400> 146
gtcgaccaaa gaaacagctt c 21

<210> 147
<211> 34
<212> DNA
<213> Artificial sequence

<220>
<223> YGR106C-F

<400> 147
ggccattatg gccaaaatgg tgttcggtca gctg 34

<210> 148
<211> 22
<212> DNA
<213> Artificial sequence

<220>
<223> YGR106C-R

<400> 148
ccaacgcacc atatgtgata tc 22

<210> 149
<211> 35
<212> DNA
<213> Artificial sequence

<220>
<223> YDR077W-F

<400> 149
ggccattatg gccaaaatga aattatcaac tgtcc 35

<210> 150
<211> 19
<212> DNA
<213> Artificial sequence

<220>
<223> YDR077W-R

<400> 150
taacatagca acaccagcc 19

<210> 151
<211> 35
<212> DNA

<213> Artificial sequence

<220>
<223> YIL123W-F

<400> 151
ggccattatg gccaaaatga aattctcaac tgccg 35

<210> 152
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<223> YIL123W-R

<400> 152
acagagacgg tacacccgtc 20

<210> 153
<211> 36
<212> DNA
<213> Artificial sequence

<220>
<223> YNL190W-F

<400> 153
ggccattatg gccaaaatga agttctcttc tgttac 36

<210> 154
<211> 22
<212> DNA
<213> Artificial sequence

<220>
<223> YNL190W-R

<400> 154
gcaccggcta cggcagcact ac 22

<210> 155
<211> 35
<212> DNA
<213> Artificial sequence

<220>
<223> YBR078W-F

<400> 155
ggccattatg gccaaaatgc aattcaagaa cgctt 35

<210> 156
<211> 21
<212> DNA
<213> Artificial sequence

<220>
<223> YBR078W-R

<400> 156
cagtgatgaa ccaaccgtct c 21

<210> 157
<211> 36
<212> DNA
<213> Artificial sequence

<220>
<223> YJL178C-F

<400> 157
ggccattatg gccaaaatgg tatcgaagac ttggat 36

<210> 158
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<223> YJL178C-R

<400> 158
aacggcgcta taaccgcctc 20

<210> 159
<211> 36
<212> DNA
<213> Artificial sequence

<220>
<223> YMR307W-F

<400> 159
ggccattatg gccaaaatgt tgtttaaatc cctttc 36

<210> 160
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<223> YMR307W-R

<400> 160
gcaaaaccga caccagcggc 20

<210> 161
<211> 35
<212> DNA
<213> Artificial sequence

<220>
<223> YOR247W-F

<400> 161
ggccattatg gccaaaatgc ttcaatccgt tgtct 35

<210> 162
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<223> YOR247W-R

<400> 162
actggtcgaa ttagtaatcg 20

<210> 163
<211> 37
<212> DNA
<213> Artificial sequence

<220>
<223> YJL159W-F

<400> 163
ggccattatg gccaaaatgc aatacaaaaa gactttg 37

<210> 164
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<223> YJL159W-R

<400> 164
aaatcgatag cttccaagtg 20

<210> 165
<211> 35
<212> DNA
<213> Artificial sequence

<220>
<223> YOR085W-F

<400> 165
ggccattatg gccaaaatga attggctgtt tttgg 35

<210> 166
<211> 20
<212> DNA
<213> Artificial sequence

<220>

<223> YOR085W-R

<400> 166
tttgaatggt gccgataacc 20

<210> 167
<211> 36
<212> DNA
<213> Artificial sequence

<220>
<223> YKR042W-F

<400> 167
ggccattatg gccaaaatga aattatccgc tctatt 36

<210> 168
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<223> YKR042W-R

<400> 168
gacaaagtta gcagaaccag 20

<210> 169
<211> 35
<212> DNA
<213> Artificial sequence

<220>
<223> YEL060C-F

<400> 169
ggccattatg gccaaaatga agttagaaaa tactc 35

<210> 170
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<223> YEL060C-R

<400> 170
cttgggtgaa gtaaccgatg 20

<210> 171
<211> 37
<212> DNA
<213> Artificial sequence

<220>
<223> YLR390W-A-F

<400> 171
ggccattatg gccaaaatgc gtgccaccac tttatta 37

<210> 172
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<223> YLR390W-A-R

<400> 172
aacatagcgg caacagcagc 20

<210> 173
<211> 35
<212> DNA
<213> Artificial sequence

<220>
<223> YMR251W-A-F

<400> 173
ggccattatg gccaaaatga agttatctca agttg 35

<210> 174
<211> 18
<212> DNA
<213> Artificial sequence

<220>
<223> YMR251W-A-R

<400> 174
aatcaaaaag gccaaagc 18

<210> 175
<211> 99
<212> PRT
<213> Artificial sequence

<220>
<223> TFP-40-aa

<400> 175

Met Arg Leu Ser Asn Leu Ile Ala Ser Ala Ser Leu Leu Ser Ala Ala
1 5 10 15

Thr Leu Ala Ala Pro Ala Asn His Glu His Lys Asp Lys Arg Ala Val
20 25 30

Val Thr Thr Thr Val Gln Lys Gln Thr Thr Ile Ile Val Asn Gly Ala
35 40 45

Ala Ser Thr Pro Val Ala Ala Leu Glu Glu Asn Ala Val Val Asn Ser
50 55 60

Ala Pro Ala Ala Ala Thr Ser Thr Thr Ser Ser Ala Ala Ser Val Ala
65 70 75 80

Thr Ala Ala Ala Ser Ser Ser Glu Asn Asn Ser Gln Val Ser Val Ala
85 90 95

Ala Ser Ala

<210> 176
<211> 313
<212> DNA
<213> Artificial sequence

<220>
<223> TFP-40-nt

<400> 176
ggccattacg gccaaaatgc gtctctctaa cctaattgct tctgcctctc ttttatctgc 60

tgctactctt gctgctcccg ctaaccacga acacaaggac aagcgtgctg tggtcactac 120

cactgttcaa aaacaaacca ctatcattgt taatggtgcc gcttcaactc cagttgctgc 180

tttggaagaa aatgctgttg tcaactccgc tccagctgcc gctaccagta caacatcgtc 240

tgctgcttct gtagctaccg ctgctgcttc ctcttctgag aacaactcac aagtttctgt 300

ggccgcctcg gcc 313

<210> 177
<211> 85
<212> PRT
<213> Artificial sequence

<220>
<223> TFP-50-aa

<400> 177

Met Leu Gln Ser Val Val Phe Phe Ala Leu Leu Thr Phe Ala Ser Ser
1 5 10 15

Val Ser Ala Ile Tyr Ser Asn Asn Thr Val Ser Thr Thr Thr Thr Leu
20 25 30

Ala Pro Ser Tyr Ser Leu Val Pro Gln Glu Thr Thr Ile Ser Tyr Ala
35 40 45

Asp Asp Thr Thr Thr Phe Phe Val Thr Ser Thr Val Tyr Ser Thr Ser
50 55 60

Trp Phe Thr Ser Thr Ser Ala Thr Ile Thr Asn Ala Ala Ser Ser Ser
65 70 75 80

Leu Ala Ala Ser Ala
85

<210> 178
<211> 271
<212> DNA
<213> Artificial sequence

<220>
<223> TFP-50-nt

<400> 178
ggccattacg gccaaaatgc ttcaatccgt tgtctttttc gctcttttaa ccttcgcaag 60

ttctgtgtca gcgatttatt caaacaatac tgtttctaca actaccactt tagcgcccag 120

ctactccttg gtgccccaag agactaccat atcgtacgcc gacgacacca ctacctttt 180

tgtcacctca acggtctact ccacgagctg gttcacctca acttcagcca ccattaccaa 240

tgcggcctcc tcctccctgg ccgcctcggc c 271

<210> 179
<211> 116
<212> PRT
<213> Artificial sequence

<220>
<223> TFP-51-aa

<400> 179

Met Leu Gln Ser Val Val Phe Phe Ala Leu Leu Thr Phe Ala Ser Ser
1 5 10 15

Val Ser Ala Ile Tyr Ser Asn Asn Thr Val Ser Thr Thr Thr Thr Leu
20 25 30

Ala Pro Ser Tyr Ser Leu Val Pro Gln Glu Thr Thr Ile Ser Tyr Ala
35 40 45

Asp Asp Thr Thr Thr Phe Phe Ala Thr Ser Thr Val Tyr Ser Thr Ser
50 55 60

Trp Phe Thr Ser Thr Ser Ala Thr Ile Thr Asn Ala Ala Ser Ser Ser
65 70 75 80

Leu Ser Thr Ser Ser Ala Ser Gly Ser Val Thr Pro Glu Ser Thr His
85 90 95

Glu Ile Thr Ser Thr Ser Thr Ile Thr Ser Thr Ser Leu Leu Thr Leu
100 105 110

Ala Ala Ser Ala
115

<210> 180
<211> 364
<212> DNA
<213> Artificial sequence

<220>
<223> TFP-51-nt

<400> 180
ggccattacg gccaaaatgc ttcaatccgt tgtctttttc gctcttttaa ccttcgcaag 60

ttctgtgtca gcgatttatt caaacaatac tgtttctaca actaccactt tagcgcccag 120

ctactccttg gtgccccaag agactaccat atcgtacgcc gacgacacca ctacctttt 180

tgccacctca acggtctact ccacgagctg gttcacctca acttcagcca ccattaccaa 240

tgcggcctcc tcctccttgt ccacctcttc ggcctctgga tctgtaaccc cagaatccac 300

ccatgaaatt acctccacct cgactatcac gtccacttcg ctgctaaccc tggccgcctc 360

ggcc 364

<210> 181
<211> 114
<212> PRT
<213> Artificial sequence

<220>
<223> TFP-57-aa

<400> 181

Met Phe Asn Arg Phe Asn Lys Leu Gln Ala Ala Leu Ala Leu Val Leu
1 5 10 15

Tyr Ser Gln Ser Ala Leu Gly Gln Tyr Tyr Thr Asn Ser Ser Ser Ile
20 25 30

Ala Ser Asn Ser Ser Thr Ala Val Ser Ser Thr Ser Ser Gly Ser Val
35 40 45

Ser Ile Ser Ser Ser Ile Glu Leu Thr Ser Ser Thr Ser Asp Val Ser

50 55 60

Ser Ser Leu Thr Glu Leu Thr Ser Ser Ser Thr Glu Val Ser Ser Ser
65 70 75 80

Ile Ala Pro Ser Thr Ser Ser Ser Glu Val Ser Ser Ser Ile Thr Ser
85 90 95

Ser Gly Ser Ser Val Ser Gly Ser Ser Ser Ile Thr Ser Leu Ala Ala
100 105 110

Ser Ala

<210> 182
<211> 358
<212> DNA
<213> Artificial sequence

<220>
<223> TFP-57-nt

<400> 182
ggccattacg gccaaaatgt tcaatcgctt taataaactt caagccgctt tggctttggt 60

cctttactcc caaagtgcat tgggccaata ttataccaac agttcctcaa tcgctagtaa 120

cagctccacc gccgtttcgt caacttcatc aggttccgtt tccatcagta gttctattga 180

gttgacctca tctacttctg atgtctcgag ctctctcact gagttaacgt catcctccac 240

cgaagtctcg agctccattg ctccatcaac ctcgtcctct gaagtctcga gctctattac 300

ttcatcaggc tcttcagtct ccggctcatc ttctattact tccctggccg cctcggcc 358

<210> 183
<211> 199
<212> PRT
<213> Artificial sequence

<220>
<223> TFP-58-aa

<400> 183

Met Phe Asn Arg Phe Asn Lys Phe Gln Ala Ala Val Ala Leu Ala Leu
1 5 10 15

Leu Ser Arg Gly Ala Leu Gly Asp Ser Tyr Thr Asn Ser Thr Ser Ser
20 25 30

Ala Asp Leu Ser Ser Ile Thr Ser Val Ser Ser Ala Ser Ala Ser Ala
35 40 45

Thr Ala Ser Asp Ser Leu Ser Ser Ser Asp Gly Thr Val Tyr Leu Pro
50 55 60

Ser Thr Thr Ile Ser Gly Asp Leu Thr Val Thr Gly Lys Val Ile Ala
65 70 75 80

Thr Glu Ala Val Glu Val Ala Ala Gly Gly Lys Leu Thr Leu Leu Asp
85 90 95

Gly Glu Lys Tyr Val Phe Ser Ser Asp Leu Lys Val His Gly Asp Leu
100 105 110

Val Val Glu Lys Ser Glu Ala Ser Tyr Glu Gly Thr Ala Phe Asp Val
115 120 125

Ser Gly Glu Thr Phe Glu Val Ser Gly Asn Phe Ser Ala Glu Glu Thr
130 135 140

Gly Ala Val Ser Ala Ser Ile Tyr Ser Phe Thr Pro Ser Ser Phe Lys
145 150 155 160

Ser Ser Gly Asp Ile Ser Leu Ser Leu Ser Lys Ala Lys Lys Gly Glu
165 170 175

Val Thr Phe Ser Pro Tyr Ser Asn Ala Gly Thr Phe Ser Leu Ser Asn
180 185 190

Ala Ile Leu Ala Ala Ser Ala
195

<210> 184
<211> 613
<212> DNA
<213> Artificial sequence

<220>
<223> TFP-58-nt

<400> 184
ggccattacg gccaaaatgt tcaatcgttt taacaaattc caagctgctg tcgctttggc 60

cctactctct cgcggcgctc tcggtgactc ttacaccaat agcacctcct ccgcagactt 120

gagttctatc acttccgtct cgtcagctag tgcaagtgcc accgcttccg actcactttc 180

ttccagtgac ggtaccgttt atttgccatc cacaacaatt agcggtgatc tcacagttac 240

tggtaaagta attgcaaccg aggccgtgga agtcgctgcc ggtggtaagt tgactttact 300

tgacggtgaa aaatacgtct tctcatctga tctaaaagtt cacggtgatt tggttgtcga 360

aaagtctgaa gcaagctacg aaggtaccgc cttcgacgtt tctggtgaga cttttgaagt 420

ttccggtaac ttcagtgctg aagaaactgg cgctgtctcc gcatctatct attcattcac 480

acctagctcg ttcaagagca gcggtgacat ttctttgagt ttgtcaaagg ccaagaaggg 540

tgaagtcacc ttttctccat actctaacgc tggtaccttt tctttgtcaa atgctattct 600

ggccgcctcg gcc 613

<210> 185
<211> 55
<212> PRT
<213> Artificial sequence

<220>
<223> TFP-59-aa

<400> 185

Met Asn Trp Leu Phe Leu Val Ser Leu Val Phe Phe Cys Gly Val Ser
1 5 10 15

Thr His Pro Ala Leu Ala Met Ser Ser Asn Arg Leu Leu Lys Leu Ala
20 25 30

Asn Lys Ser Pro Lys Lys Ile Ile Pro Leu Lys Asp Ser Ser Phe Glu
35 40 45

Asn Ile Leu Ala Ala Ser Ala
50 55

<210> 186
<211> 181
<212> DNA
<213> Artificial sequence

<220>
<223> TFP-59-nt

<400> 186
ggccattacg gccaaaatga attggctgtt tttggtctcg ctggttttct tctgcggcgt 60

gtcaacccat cctgccctgg caatgtccag caacagacta ctaaagctgg ctaataaatc 120

tcccaagaaa attatacctc tgaaggactc aagttttgaa aacatcctgg ccgcctcggc 180

c 181

<210> 187
<211> 34
<212> DNA
<213> Artificial sequence

<220>
<223> T1-F

<400> 187
ggccattacg gccaaaatgt tcaatcgttt taac 34

<210> 188
<211> 37
<212> DNA
<213> Artificial sequence

<220>
<223> T1-R

<400> 188
ttgtagtgtt gactggagca ccgagagcgc cgcgaga 37

<210> 189
<211> 35
<212> DNA
<213> Artificial sequence

<220>
<223> T2-F

<400> 189
ggccattacg gccaaaatga cgccctatgc agtag 35

<210> 190
<211> 37
<212> DNA
<213> Artificial sequence

<220>
<223> T2-R

<400> 190
ttgtagtgtt gactggagct gcgctcactg ttacaat 37

<210> 191
<211> 34
<212> DNA
<213> Artificial sequence

<220>
<223> T3-F

<400> 191
ggccattacg gccaaaatgc aattcaaaaa cgtc 34

<210> 192
<211> 37
<212> DNA
<213> Artificial sequence

<220>
<223> T3-R

<400> 192
ttgtagtgtt gactggagca gcagaagcag tggcgga 37

<210> 193
<211> 34
<212> DNA
<213> Artificial sequence

<220>
<223> T4-F

<400> 193
ggccattacg gccaaaatga gatttgcaga attc 34

<210> 194
<211> 38
<212> DNA
<213> Artificial sequence

<220>
<223> T4-R

<400> 194
ttgtagtgtt gactggagca gccatccccc cgcctaac 38

<210> 195
<211> 18
<212> DNA
<213> Artificial sequence

<220>
<223> MF-pro-F

<400> 195
gctccagtca acactaca 18

<210> 196
<211> 36
<212> DNA
<213> Artificial sequence

<220>
<223> MF-R

<400> 196
ggccgaggcg gccgataccc cttcttcttt agcagc 36

<210> 197
<211> 34
<212> DNA
<213> Artificial sequence

<220>

<223> MF-Pre-F

<400> 197
ggccattacg gccaaaatgg tatcgaagac ttgg 34

<210> 198
<211> 36
<212> DNA
<213> Artificial sequence

<220>
<223> KR-IGF-F

<400> 198
ctcgccttag ataaaagagg accggagacg ctctgc 36

<210> 199
<211> 36
<212> DNA
<213> Artificial sequence

<220>
<223> IGF-R

<400> 199
cactccgttc aagtcgactc aagctgactt ggcagg 36

<210> 200
<211> 88
<212> PRT
<213> Artificial sequence

<220>
<223> TFP-5-aa

<400> 200

Met Phe Asn Arg Phe Asn Lys Phe Gln Ala Ala Val Ala Leu Ala Leu
1 5 10 15

Leu Ser Arg Gly Ala Leu Gly Ala Pro Val Asn Thr Thr Thr Glu Asp
20 25 30

Glu Thr Ala Gln Ile Pro Ala Glu Ala Val Ile Gly Tyr Leu Asp Leu
35 40 45

Glu Gly Asp Phe Asp Val Ala Val Leu Pro Phe Ser Asn Ser Thr Asn
50 55 60

Asn Gly Leu Leu Phe Ile Asn Thr Thr Ile Ala Ser Ile Ala Ala Lys
65 70 75 80

Glu Glu Gly Val Ala Ala Ser Ala

85

<210> 201
<211> 280
<212> DNA
<213> Artificial sequence

<220>
<223> TFP-5-NT

<400> 201
ggccattacg gccaaaatgt tcaatcgttt taacaaattc caagctgctg tcgctttggc 60

cctactctct cgcggcgctc tcggtgctcc agtcaacact acaacagaag atgaaacggc 120

acaaattccg gctgaagctg tcatcggtta cttagattta gaaggggatt tcgatgttgc 180

tgttttgcca ttttccaaca gcacaaataa cgggttattg tttataaata ctactattgc 240

cagcattgct gctaaagaag aaggggtggc cgcctcggcc 280

<210> 202
<211> 84
<212> PRT
<213> Artificial sequence

<220>
<223> TFP-6-aa

<400> 202

Met Thr Pro Tyr Ala Val Ala Ile Thr Val Ala Leu Leu Ile Val Thr
1 5 10 15

Val Ser Ala Ala Pro Val Asn Thr Thr Thr Glu Asp Glu Thr Ala Gln
20 25 30

Ile Pro Ala Glu Ala Val Ile Gly Tyr Leu Asp Leu Glu Gly Asp Phe
35 40 45

Asp Val Ala Val Leu Pro Phe Ser Asn Ser Thr Asn Asn Gly Leu Leu
50 55 60

Phe Ile Asn Thr Thr Ile Ala Ser Ile Ala Ala Lys Glu Glu Gly Val
65 70 75 80

Ala Ala Ser Ala

<210> 203
<211> 268
<212> DNA
<213> Artificial sequence

<220>
<223> TFP-6-NT

<400> 203
ggccattacg gccaaaatga cgccctatgc agtagcaatt accgtggcct tactaattgt 60

aacagtgagc gcagctccag tcaacactac aacagaagat gaaacggcac aaattccggc 120

tgaagctgtc atcggttact tagatttaga aggggatttc gatgttgctg ttttgccatt 180

ttccaacagc acaaataacg ggttattgtt tataaatact actattgcca gcattgctgc 240

taaagaagaa ggggtggccg cctcggcc 268

<210> 204
<211> 86
<212> PRT
<213> Artificial sequence

<220>
<223> TFP-7-aa

<400> 204

Met Gln Phe Lys Asn Val Ala Leu Ala Ala Ser Val Ala Ala Leu Ser
1 5 10 15

Ala Thr Ala Ser Ala Ala Pro Val Asn Thr Thr Thr Glu Asp Glu Thr
20 25 30

Ala Gln Ile Pro Ala Glu Ala Val Ile Gly Tyr Leu Asp Leu Glu Gly
35 40 45

Asp Phe Asp Val Ala Val Leu Pro Phe Ser Asn Ser Thr Asn Asn Gly
50 55 60

Leu Leu Phe Ile Asn Thr Thr Ile Ala Ser Ile Ala Ala Lys Glu Glu
65 70 75 80

Gly Val Ala Ala Ser Ala
85

<210> 205
<211> 274
<212> DNA
<213> Artificial sequence

<220>
<223> TFP-7-NT

<400> 205
ggccattacg gccaaaatgc aattcaaaaa cgtcgcccta gctgcctccg ttgctgctct 60

atccgccact gcttctgctg ctccagtcaa cactacaaca gaagatgaaa cggcacaaat 120

tccggctgaa gctgtcatcg gttacttaga tttagaaggg gatttcgatg ttgctgtttt 180

gccattttcc aacagcacaa ataacgggtt attgtttata aatactacta ttgccagcat 240

tgctgctaaa gaagaagggg tggccgcctc ggcc 274

<210> 206
<211> 83
<212> PRT
<213> Artificial sequence

<220>
<223> TFP-8-aa

<400> 206

Met Arg Phe Ala Glu Phe Leu Val Val Phe Ala Thr Leu Gly Gly Gly
1 5 10 15

Met Ala Ala Pro Val Asn Thr Thr Thr Glu Asp Glu Thr Ala Gln Ile
20 25 30

Pro Ala Glu Ala Val Ile Gly Tyr Leu Asp Leu Glu Gly Asp Phe Asp
35 40 45

Val Ala Val Leu Ser Phe Ser Asn Ser Thr Asn Asn Gly Leu Leu Phe
50 55 60

Ile Asn Thr Thr Ile Ala Ser Ile Ala Ala Lys Glu Glu Gly Val Ala
65 70 75 80

Ala Ser Ala

<210> 207
<211> 265
<212> DNA
<213> Artificial sequence

<220>
<223> TFP-8-NT

<400> 207
ggccattacg gccaaaatga gatttgcaga attcttggtg gtatttgcca cgttaggcgg 60

ggggatggct gctccagtca acactacaac agaagatgaa acggcacaaa ttccggctga 120

agctgtcatc ggttacttag atttagaagg ggatttcgat gttgctgttt tgtcattttc 180

caacagcaca aataacgggt tattgtttat aaatactact attgccagca ttgctgctaa 240

agaagaaggg gtggccgcct cggcc 265

<210> 208
<211> 84
<212> PRT
<213> Artificial sequence

<220>
<223> TFP-32-aa

<400> 208

Met Arg Phe Pro Ser Ile Phe Thr Ala Val Leu Phe Ala Ala Ser Ser
1 5 10 15

Ala Leu Ala Ala Pro Val Asn Thr Thr Thr Glu Asp Glu Thr Ala Gln
20 25 30

Ile Pro Ala Glu Ala Val Ile Gly Tyr Leu Asp Leu Glu Gly Asp Phe
35 40 45

Asp Val Ala Val Leu Pro Phe Ser Asn Ser Thr Asn Asn Gly Leu Leu
50 55 60

Phe Ile Asn Thr Thr Ile Ala Ser Ile Ala Ala Lys Glu Glu Gly Val
65 70 75 80

Ala Ala Ser Ala

<210> 209
<211> 268
<212> DNA
<213> Artificial sequence

<220>
<223> TFP-32-NT

<400> 209
ggccattacg gccaaaatga gatttccttc aatttttact gcagttttat tcgcagcatc 60

ctccgcatta gctgctccag tcaacactac aacagaagat gaaacggcac aaattccggc 120

tgaagctgtc atcggttact tagatttaga aggggatttc gatgttgctg ttttgccatt 180

ttccaacagc acaaataacg ggttattgtt tataaatact actattgcca gcattgctgc 240

taaagaagaa ggggtggccg cctcggcc 268

<210> 210
<211> 36
<212> DNA
<213> Artificial sequence

<220>
<223> KR-hP10-F

<400> 210
ctcgccttag ataaaagagc tattaagaaa gcccac 36

<210> 211
<211> 35
<212> DNA
<213> Artificial sequence

<220>
<223> hP10-Sal-R

<400> 211
ctccgttcaa gtcgacttaa tgtcctggga agagg 35

<210> 212
<211> 37
<212> DNA
<213> Artificial sequence

<220>
<223> KR-IL32g-F

<400> 212
ctcgccttag ataaaagaat gtgcttcccg aaggtcc 37

<210> 213
<211> 36
<212> DNA
<213> Artificial sequence

<220>
<223> IL32g-SalI-R

<400> 213
cactccgttc aagtcgactc attttgagga ttgggg 36

<210> 214
<211> 4
<212> PRT
<213> Artificial sequence

<220>
<223> kex2p protease recognition sequence

<400> 214

Leu Asp Lys Arg
1

<210> 215
<211> 4
<212> PRT
<213> Artificial sequence

<220>
<223> Factor Xa recognition sequence

<400> 215

Ile Glu Gly Arg
1

<210> 216
<211> 4
<212> PRT
<213> Artificial sequence

<220>
<223> subtilisin recognition sequence

<400> 216

Ala Ala His Tyr
1

<210> 217
<211> 7
<212> PRT
<213> Artificial sequence

<220>
<223> tobacco etch virus recognition sequence

<400> 217

Glu Asn Leu Tyr Phe Gln Gly
1 5

<210> 218
<211> 7
<212> PRT
<213> Artificial sequence

<220>
<223> thrombin recognition sequence

<400> 218

Glu Asn Leu Tyr Phe Gln Gly
1 5

<210> 219
<211> 105
<212> PRT
<213> Artificial sequence

<220>
<223> TFP-1-aa

<400> 219

```
Met Phe Asn Arg Phe Asn Lys Phe Gln Ala Ala Val Ala Leu Ala Leu
1               5                  10                  15

Leu Ser Arg Gly Ala Leu Gly Asp Ser Tyr Thr Asn Ala Thr Ser Ser
            20                  25                  30

Ala Asp Leu Ser Ser Ile Thr Ser Val Ser Ser Ala Ser Ala Ser Ala
        35                  40                  45

Thr Ala Ser Asp Ser Leu Ser Ser Ser Asp Gly Thr Val Tyr Leu Pro
    50                  55                  60

Ser Thr Thr Ile Ser Gly Asp Leu Thr Val Thr Gly Lys Val Ile Ala
65                  70                  75                  80

Thr Glu Ala Val Glu Val Ala Ala Gly Gly Lys Leu Thr Leu Leu Asp
                85                  90                  95

Gly Glu Lys Tyr Val Phe Ser Ser Asp
            100                 105
```

<210> 220
<211> 430
<212> DNA
<213> Artificial sequence

<220>
<223> TFP-1-NT

<400> 220
gatcgtcata ttcactcttg ttctcataat agcagtccaa gttttcatct ttgcaagctt 60

tactatttct ttctttttat tggtaaactc tcgcccatta caaaaaaaaa agagatgttc 120

aatcgtttta acaaattcca agctgctgtc gctttggccc tactctctcg cggcgctctc 180

ggtgactctt acaccaatag cacctcctcc gcagacttga gttctatcac ttccgtctcg 240

tcagctagtg caagtgccac cgcttccgac tcactttctt ccagtgacgg taccgtttat 300

ttgccatcca caacaattag cggtgatctc acagttactg gtaaagtaat tgcaaccgag 360

gccgtggaag tcgctgccgg tggtaagttg actttacttg acggtgaaaa atacgtcttc 420

tcatctgatc 430

<210> 221
<211> 117
<212> PRT
<213> Artificial sequence

<220>

<223> TFP-2-aa

<400> 221

Met Thr Pro Tyr Ala Val Ala Ile Thr Val Ala Leu Leu Ile Val Thr
1 5 10 15

Val Ser Ala Leu Gln Val Asn Asn Ser Cys Val Ala Phe Pro Pro Ser
20 25 30

Asn Leu Arg Gly Lys Asn Gly Asp Gly Thr Asn Glu Gln Tyr Ala Thr
35 40 45

Ala Leu Leu Ser Ile Pro Trp Asn Gly Pro Pro Glu Ser Leu Arg Asp
50 55 60

Ile Asn Leu Ile Glu Leu Glu Pro Gln Val Ala Leu Tyr Leu Leu Glu
65 70 75 80

Asn Tyr Ile Asn His Tyr Tyr Asn Thr Thr Arg Asp Asn Lys Cys Pro
85 90 95

Asn Asn His Tyr Leu Met Gly Gly Gln Leu Gly Ser Ser Ser Asp Asn
100 105 110

Arg Ser Leu Asn Asp
115

<210> 222
<211> 424
<212> DNA
<213> Artificial sequence

<220>
<223> TFP-2-NT

<400> 222
gatctcattg gattcaagag aaagaaactc tatactggcg ccaaattagc agtgtcaaat 60

ttcgaaaagg tgatgacgcc ctatgcagta gcaattaccg tggccttact aattgtaaca 120

gtgagcgcac tccaggtcac aattcatgtg tcgcttttcc gccaatcaaa tctcagaggc 180

aaaaatggag acggtactaa tgaacagtat gcaactgcac tactttctat tccctggaat 240

ggacctcctg agtcattgag ggatattaat cttattgaac tcgaaccgca agttgcactc 300

tatttgctcg aaaattatat taaccattac tacaacacca caagagacaa taagtgccct 360

aataaccact acctaatggg agggcagttg ggtagctcat cggataatag gagtttgaac 420

gatc 424

<210> 223
<211> 104
<212> PRT
<213> Artificial sequence

<220>
<223> TFP-3-aa

<400> 223

Met Gln Phe Lys Asn Val Ala Leu Ala Ala Ser Val Ala Ala Leu Ser
1 5 10 15

Ala Thr Ala Ser Ala Glu Gly Tyr Thr Pro Gly Glu Pro Trp Ser Thr
20 25 30

Leu Thr Pro Thr Gly Ser Ile Ser Cys Gly Ala Ala Glu Tyr Thr Thr
35 40 45

Thr Phe Gly Ile Ala Val Gln Ala Ile Thr Ser Ser Lys Ala Lys Arg
50 55 60

Asp Val Ile Ser Gln Ile Gly Asp Gly Gln Val Gln Ala Thr Ser Ala
65 70 75 80

Ala Thr Ala Gln Ala Thr Asp Ser Gln Ala Gln Ala Thr Thr Thr Ala
85 90 95

Thr Pro Thr Ser Ser Glu Lys Ile
100

<210> 224
<211> 642
<212> DNA
<213> Artificial sequence

<220>
<223> TFP-3-NT

<400> 224
gatcccgcct agcccttcca gcttttcttt ttccccttt gctacggtcg agacacggtc 60

gcccaaaaga aacgggtcag cgtgtactgc gccaaaaaaa ttcgcgccga tttaagctaa 120

acgtccacaa acaaaaacaa aaataagaaa taggttgaca gtgggtgaaa aattctcgaa 180

ggtttcatct ccaaacagtc agtatataag tattcgggaa agagagccaa tctatcttgt 240

ggtgggtcta tcttaacctt ctctttttgg cagtagtaat tgtaaatcaa gacacataaa 300

actatttcac tcgctaaact tacatctaaa atgcaattca aaaacgtcgc cctagctgcc 360

tccgttgctg ctctatccgc cactgcttct gctgaaggtt acactccagg tgaaccatgg 420

tccaccttaa ccccaaccgg ctccatctct tgtggtgctg ccgaatacac taccaccttt 480

ggtattgctg ttcaagctat tacctcttca aaagctaaga gagacgttat ctctcaaatt 540

ggtgacggtc aagtccaagc cacttctgct gctactgctc aagccaccga tagtcaagcc 600

caagctacta ctaccgctac cccaaccagc tccgaaaaga tc 642

<210> 225
<211> 50
<212> PRT
<213> Artificial sequence

<220>
<223> TFP-4-aa

<400> 225

Met Arg Phe Ala Glu Phe Leu Val Val Phe Ala Thr Leu Gly Gly Gly
1 5 10 15

Met Ala Ala Pro Val Glu Ser Leu Ala Gly Thr Gln Arg Tyr Leu Val
20 25 30

Gln Met Lys Glu Arg Phe Thr Thr Glu Lys Leu Cys Ala Leu Asp Asp
35 40 45

Lys Ile
50

<210> 226
<211> 179
<212> DNA
<213> Artificial sequence

<220>
<223> TFP-4-NT

<400> 226
gatccgcttt ttattgcttt gctttgctaa tgagatttgc agaattcttg gtggtatttg 60

ccacgttagg cgggggggatg gctgcaccgg ttgagtctct ggccgggacc caacggtatc 120

tggtgcaaat gaaggagcgg ttcaccacag agaagctgtg tgctttggac gacaagatc 179

<210> 227
<211> 35
<212> DNA
<213> Artificial sequence

<220>
<223> PpTFP1-F

<400> 227

agtggccatt acggccaaaa tgcaattcaa cagtg 35

<210> 228
<211> 39
<212> DNA
<213> Artificial sequence

<220>
<223> PpTFP1-R

<400> 228
tagggccgag gcggccagtg tggccgatgg gtcccattg 39

<210> 229
<211> 35
<212> DNA
<213> Artificial sequence

<220>
<223> PpTFP2-F

<400> 229
agtggccatt acggccaaaa tgcaattctc tatcg 35

<210> 230
<211> 38
<212> DNA
<213> Artificial sequence

<220>
<223> PpTFP2-R

<400> 230
tagggccgag gcggccagtg gggtggagtg ggtggttg 38

<210> 231
<211> 37
<212> DNA
<213> Artificial sequence

<220>
<223> PpTFP3-F

<400> 231
agtggccatt acggccaaaa tgaagttttc cactgcg 37

<210> 232
<211> 38
<212> DNA
<213> Artificial sequence

<220>
<223> PpTFP3-R

<400> 232
tagggccgag gcggccaggg tagtggtagg atctggag 38

<210> 233
<211> 35
<212> DNA
<213> Artificial sequence

<220>
<223> PpTFP4-F

<400> 233
agtggccatt acggccaaaa tgcaatacag atctc 35